ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# THE AXIOMS AND ALGEBRA OF INTUITIVE PROBABILITY

BY B. O. KOOPMAN

(Received June 1, 1939)

## 1. The Idea of Probability

The intuitive thesis in probability holds that both in its meaning and in the laws which it obeys, probability derives directly from the intuition, and is prior to objective experience;[1] it holds that it is for experience to be interpreted in terms of probability and not for probability to be interpreted in terms of experience; and it holds that all the so-called objective definitions of probability depend for their effective application to concrete cases upon their translation into the terms of intuitive probability. In its purest form it stems from the aphorism: *knowledge is possible, while certainty is not*; for it sees in the resolution of this antinomy—it sees in the varying "probability" in which the articles of knowledge are held—its very conceptual germ.

Clearly the vindication of a thesis such as this falls outside the domain of mathematician or physicist, and belongs to the philosopher: the question at issue is one of epistemology.[2] And our function here is to undertake neither its defense nor its refutation, but solely to attempt the axiomatic formulation of the idea of intuitive probability, and the derivation of the classical mathematical theory of probability from this. That such a task is necessary at the present time appears from the fact that all the axiomatic treatments of intuitive probability current in the literature take as their starting point a number (usually between 0 and 1) corresponding to the "degree of rational belief" or "credibility" of the eventuality in question. Now we hold that such a number is in no wise a self-evident concomitant with or expression of the primordial intuition of probability, but rather a mathematical construct derived from the latter under very special conditions and as the result of a fairly complicated process implicitly based on many of the very intuitive assumptions which we are endeavouring to axiomatize: There is, in short, what appears to us to be a

[1] "(Objective) experience" is used here practically as the equivalent of "laboratory experiment" in the narrow objective sense of the word, and excludes introspective or "subjective experience." Moreover, it is a question here of *rational* derivation from experience, and the *a priori* view of probability is quite consistent with the idea that probability as well as logic may be derived by race experience through the process of evolution. In spite of terminological similarities, there is no question here of Kantean Transcendentalism.

[2] The philosophical controversy here involved is very old and is still unsettled. Without attempting to give general references to the literature on this point we may cite the Introduction and First Chapter of J. M. Keynes' *A Treatise on Probability* (London, 1921), in which the thesis of intuitive probability is eloquently urged.

serious rational lacuna between the primal intuition of probability, and that branch of the theory of measure which passes conventionally under the name of probability. Moreover, this assumption commits one to too great precision, leading either to absurdities, or else to the undue restriction of the field of applicability of the idea of probability.[3]

The fundamental viewpoint of the present work is that the primal intuition of probability expresses itself in a (partial) ordering of eventualities: A certain individual at a certain moment considers the propositions $a, b, h, k$, the meanings of which he apprehends, and which he considers to be determinate (i.e., either true or false) without (in general) knowing whether they are true or false (for the sharpening of these ideas, see §2). Then the phrase

> *"a on the presumption that h is true is equally or less probable than b on the presumption that k is true"*

conveys a precise meaning to his intuition—and this, utterly disregarding whether he accepts what it says, rejects it, or is non-committal. This is, as we see it, a first essential in the thesis of intuitive probability, and contains the ultimate answer to the question of the meaning of the notion of probability. Let us write the quoted phrase in either of the symbolic forms

$$a/h \prec b/k, \qquad b/k \succ a/h$$

and call such an expression a *comparison in probability.*[4] It will be convenient to call such symbols as $a/h$ or $b/k$ *eventualities*, the propositions $a$ and $b$, the *contingencies* in the eventualities, and the propositions $h$ and $k$, the *presumptions* in the eventualities. And we may read $a/h \prec b/k$: $a$ on $h$ *infraprobable* $b$ on $k$ ($\succ$ being read *supraprobable*).

The position of acceptance, rejection, or doubt taken with regard to $a/h \prec b/k$ may vary from individual to individual and from moment to moment with the same individual, depending as it will upon his entire state of mind. Never-

[3] We have in mind, on the one hand, events of such elaborate and individual nature as to defy numerical evaluation of probability, yet which can often be definitely compared as to their likelihood of occurrence; and on the other hand, instances such as the following: Let 0 denote an impossible event, let $A$ and $B$ consist of a certain normally distributed variate $x$ being found to have exactly the value $a$ and $b$ respectively, $b$ being much closer to the central value than $a$, so that for the probability densities: $\varphi(a) < \varphi(b)$. Then as far as the numerical probabilities go, $p(0) = p(A) = p(B) = 0$, but it is rational to place them in the order of likelihood: $0 < A < B$.

But fundamentally our object is not so much the treatment of such cases, as to secure an adequate conceptual framework for the idea of probability.

[4] The notation $a/h$ is suggested by that of Keynes (l.c.); yet we are using it in quite a different sense and in a manner much closer to the current notation for the quotient of a ring by an ideal (the ideal $(\sim h)$ cf. §2). For to begin with, Keynes regards $a/h$ as a number, or at least susceptible of many of the rules of ordinary arithmetic, whereas for us it is but an ordered pair of propositions, or, as we shall see later, a remainder class in a Boolean ring. And secondly, the *logical type* of proposition which we denote by $a$, $h$, etc., is far different from what Keynes appears to understand by such letters.

theless, if a certain individual at a certain moment subscribes to both of the following: $a/h \prec b/k$, $b/k \prec c/l$, he *must* at the same moment subscribe to $a/h \prec c/l$, and also he must subscribe to each of $\sim a/h \succ \sim b/k$, $\sim b/k \succ \sim c/l$. For otherwise he will be in a condition of mental incoherence of the same nature as logical inconsistency. We thus come to a second essential in the thesis of intuitive probability, namely, that its comparisons obey fixed laws which, like the laws of intuitive logic to which they are akin, are axiomatic, and owe their authority to man's awareness of his own rational processes.[5]

The circumstance that notwithstanding the "subjectivity" with which comparisons in probability may be made, they are none the less fixed both as regards their meaning and their laws, is the very fact which makes a rational mathematical theory of probability possible. *It is the unique function of such a theory to develop the rules for the derivation of comparisons in probability from other comparisons in probability previously given.*

When the intuition exerts itself upon a situation involving uncertain events, it is capable of performing two tasks: firstly, it may assert comparisons in probability; and secondly, it may state laws of consistency governing such comparisons. There is between these two operations of the intuition an essential cleft; and the failure to realize this has been a most frequent cause of confusion. From the laws of consistency a theory may be constructed (and this is our present object) with the aid of which, comparisons in probability (and indeed, values of numerical probability) can be deduced from others which are hypothesized: but beyond this, the laws of consistency do not allow one to go. This naturally raises the question as to whether there are any general principles which determine (non-trivial) comparisons in probability *ab ovo*,—without taking any such comparisons as starting point.

One might consider a concrete case: Two wayfarers are on a road which, as they know, leads to their destination. After a while, to their surprize, the road divides in two, and as they are ignorant of their orientation, they do not know which way to turn. The first wayfarer asserts that the right-hand turn is equally or less probable as the way to their destination than the left; but the second disagrees with him. *Is there any general principle which will enable either one rationally to compel the other to accept his position?* We mean a principle which does not itself implicitly assume a comparison in probability (such as a statement of "irrelevancy") to have been granted and which can be applied to the present case without begging the question; a principle which, finally,

[5] There is a body of opinion at the present time which regards the laws of probability, and even the laws of logic, as subject to objective experimental proof or disproof, or at least as requiring possible alteration in the light of the evolution of physical theory. This school strengthens its case by the claim that certain facts in the quantum theory are at variance with the "laws of probability" (or indeed, of logic). While we shall treat this question in detail elsewhere, we may say here that we have examined with all logical minuteness every alleged case which we could discover, and have found in every case that the physical circumstances of the application of the laws of probability, etc., were the things that were altered, but never the laws themselves.

genuinely pertains to the general abstract theory of probability, and not to some specialized branch of experimental science.

While an authoritative answer to this question would go beyond the scope of the present paper, involving as it would a critique of the existing state of the foundations of probability *in toto*, we may nevertheless venture an expression of our personal view: We have never seen formulated nor been able ourselves to formulate such a principle. And we have come to the position that no such principle is in the nature of things susceptible of formulation, that we are indeed at the very threshold of the problem of the conditions of the possibility of knowledge, and that the difficulties here met with are those which must in the nature of things always be encountered when an attempt is made to give a mathematical or physical solution to a metaphysical problem.

## 2. Elementary Notions of Logic

In the theory of probability one is constantly occupied with what may be called determinate concrete propositions, that is to say, specific assertions of a physical or biological character, the truth or falsity of which is regarded as verifiable on the performance of an appropriate experiment. Such propositions and such only shall in the present paper be symbolized by small Latin letters or by finite clusters of such letters united with the aid of the logical constants $(\sim \cdot \vee)$.[6] For example, the proposition "it will rain (at stated time and place)" would be denoted by a letter such as $a$, "it will rain or snow (at stated time and place," by $c$ or by $a \vee b$, etc. Now all propositions of this sort are conceived as *contemplated* propositions (cf. Frege and Russell) and not as mere symbols for 0 or 1:[7] we may talk about $a$ and $a \vee b$ without automatically asserting them. Altogether different shall be our conventions regarding the use of the logical constants $(= \subset)$,[8] which shall be regarded as having predicative power. Thus "$c = (a \vee b)$" is an *asserted* proposition and shall never in the present text be symbolized by a single letter, nor shall assertions like it be united by means of logical constants: no formula shall contain more than one of the signs $(= \subset)$.

To sum up: We are dealing exclusively with intuitive logic, are employing the logical constants solely as abbreviations for notions which are regarded as intuitively evident and as obeying intuitively evident laws, and we are symbolizing by letters only propositions of the lowest "logical type."

---

[6] As usual, $\sim$ denotes negation, $\cdot$ (or mere juxtaposition) denotes conjunction or logical product, and $\vee$ denotes disjunction or logical sum.

[7] The identically false proposition is denoted by 0 and is regarded as coinciding with every $a \sim a$, the identically true proposition is denoted by 1 and is regarded as coinciding with every $a \vee \sim a$. Thus all the Boolean algebras here considered are regarded as having identical units.

[8] The symbol for identity $=$ and material implication $\subset$ ($a \subset b$ meaning "$a$ implies $b$") are used here rather than $\equiv$ and $\supset$ in order to bring out the parallel with the mathematical theories of sets and algebras.

Parentheses shall be used in the self-evident manner of classical mathematics; but they are economized by the conventions concerning the binding strength of symbols, according to which: (i) $\sim$ is linked most strongly to the succeeding nearest letter, so that $\sim ab$ or $\sim a \cdot b$ each mean $(\sim a) \cdot b$, $\sim a \vee b$ means $(\sim a) \vee b$, and $\sim a = b$ and $\sim a \subset b$ mean $(\sim a) = b$ and $(\sim a) \subset b$ respectively; (ii) $\cdot$ (or the absence of symbol) binds the letters on either side of it more strongly than $\vee$ so that $a \cdot b \vee c$ or $ab \vee c$ means $(ab) \vee c$. In this connection it may be remarked that in accordance with our fundamental convention, such expressions as $a \vee (b \subset c)$ or $\sim(a \subset b)$ would have no use.

The knowledge of the laws which the symbols $(\sim \cdot \vee = \subset)$ obey, or formal statement of the laws of intuitive logic, is here regarded as aquired and the reader is assumed to be familiar with the elementary facts of Boolean algebra.[9] We shall make explicit mention only of the following definitions and theorems, on account of the prominent rôle which they are to play in our theory of probability. The reader may easily furnish the proofs by analogy with the theorems on homomorphism of groups or rings, and the quotient group or remainder ring in classical algebra. Detailed proofs are furnished in the paper of M. H. Stone cited above.

DEFINITION. *By a homomorphism $\mathfrak{A} \to \mathfrak{A}'$ from a Boolean ring $\mathfrak{A}$ to a Boolean ring $\mathfrak{A}'$ is meant a functional correspondence $x' = f(x)$ whose domain is $\mathfrak{A}$ and whose range is $\mathfrak{A}'$, such that, for all $x, y \in \mathfrak{A}$,*

$$f(\sim x) = \sim f(x), \qquad f(xy) = f(x)f(y), \qquad f(x \vee y) = f(x) \vee f(y).$$

*In the case where $f(x) = f(y)$ implies $x = y$, the correspondence $\mathfrak{A} \to \mathfrak{A}'$ can be inverted (the result $\mathfrak{A}' \to \mathfrak{A}$ being shown to be a homomorphism) and the correspondence is called an isomorphism $\mathfrak{A} \leftrightarrow \mathfrak{A}'$.*

DEFINITION. *A non-empty subset $\mathfrak{u}$ of elements of a Boolean ring $\mathfrak{A}$ is said to be an ideal in $\mathfrak{A}$ when it satisfies the following conditions*:

1. *If $a \in \mathfrak{u}$ and $b \in \mathfrak{u}$, then $a \vee b \in \mathfrak{u}$*
2. *If $a \in \mathfrak{u}$ and $b \in \mathfrak{A}$ then $ab \in \mathfrak{u}$*

THEOREM. *If $\mathfrak{A} \to \mathfrak{A}'$, then the aggregate $\mathfrak{u}$ of all elements $x$ of $\mathfrak{A}$ such that $x' = f(x) = 0'$ (the zero in $\mathfrak{A}'$) constitutes an ideal in $\mathfrak{A}$. And $f(a) = f(b)$ if and only if simultaneously $a \sim b \in \mathfrak{u}$, $\sim ab \in \mathfrak{u}$.*

We shall now recall the definition of the quotient ideal $\mathfrak{A}/\mathfrak{u}$ of $\mathfrak{A}$ with respect to any one of its ideals $\mathfrak{u}$.

First, all the elements of $\mathfrak{A}$ are distributed into a system of mutually exclusive and exhaustive *remainder classes* of $\mathfrak{u}$ defined by the convention that any two elements $a$ and $b$ shall belong to the same remainder class of $\mathfrak{u}$ if and only if

[9] For general bibliographical references, see E. V. Huntington, Trans. Amer. Math. Soc., Vol. 35 (1933), pp. 274–304. For algebraic developments of Boolean rings, particularly in the light of their relation with the rings and ideals of classical abstract algebra, see M. H. Stone, *The Theory of Representations for Boolean Algebras*, Trans. Amer. Math. Soc., Vol. 40 (1936), pp. 37–111.

simultaneously $a \sim b \in \mathfrak{u}$, $\sim ab \in \mathfrak{u}$. Let the remainder class determined by the fact that it contains the element $a$ be denoted by $a/\mathfrak{u}$.

Next, if $(\sim \cdot \vee)$ applied to classes be used to denote the classes resulting from the application of the respective operator to all elements or pairs of elements of the classes in question, and if $(= \subset)$ be used to denote class identity or inclusion, it is shown that

$$\sim a/\mathfrak{u} = \sim(a/\mathfrak{u}), \qquad ab/\mathfrak{u} = (a/\mathfrak{u})(b/\mathfrak{u}), \qquad a \vee b/\mathfrak{u} = (a/\mathfrak{u}) \vee (a/\mathfrak{u}).$$

Finally, $\mathfrak{A}/\mathfrak{u}$ is defined as the class of all remainder classes in $\mathfrak{A}$ of $\mathfrak{u}$, the $(\sim \cdot \vee)$ being defined as before, $\mathfrak{A}/\mathfrak{u}$ is seen to be a Boolean ring (with units $0/\mathfrak{u}$ and $1/\mathfrak{u}$) and the correspondence $x \to x/\mathfrak{u}$ $(x \in \mathfrak{A})$ is a homomorphism: $\mathfrak{A} \to \mathfrak{A}/\mathfrak{u}$. And the following theorem is an immediate consequence.

THEOREM. *If* $\mathfrak{A} \to \mathfrak{A}'$, *and if* $\mathfrak{u}$ *is the class of all elements of* $\mathfrak{A}$ *having* $0'$ *as their image in* $\mathfrak{A}'$, *then* $\mathfrak{A}' \leftrightarrow \mathfrak{A}/\mathfrak{u}$.

Turning now to the significance of these algebraic notions for the present theory of probability, suppose that an aggregate of comparisons in probability is being envisaged, involving a set of concrete propositions $a, b, h, k$, etc. After the possible adjunction to this set of all of the finite combinations of its propositions in terms of $(\sim \cdot \vee)$—in case such combinations are not already included in the set—it becomes closed, and forms a Boolean ring $\mathfrak{A}$, containing 0, 1.

In $\mathfrak{A}$ the element $\sim h$ (assuming $\sim h \neq 1$ i.e., $h \neq 0$) generates a *principal ideal* (i.e., ideal generated by an unique element $\neq 1$), which shall be denoted by $(\sim h)$, and which is composed of the class of all elements $\sim hx$, $x \in \mathfrak{A}$. If, now, $a$ and $b$ are two elements of $\mathfrak{A}$ for which simultaneously $a \sim b \in (\sim h)$, $\sim ab \in (\sim h)$, then *a and b are logically equivalent on the assumption that h is true* in the following intuitively self-explanatory sense: since $a \sim b = \sim hc (c \in \mathfrak{A})$, to assume $h$ true is to assume $a \sim b = (a \sim b)h = h \sim hc = 0$, i.e., to rule out that $a$ is true without $b$ being true: $a \subset b$. And similarly, if $\sim ab = \sim hd$ $(d \in \mathfrak{A})$, then the assumption that $h$ is true automatically makes $\sim ab = 0$, i.e., makes $b \subset a$.

It is convenient in this work to introduce the symbol $a/h$ as simply an alternative notation for $a/(\sim h)$: in the light of the above, we become aware of the fact that the intuitive act of passing from the consideration of $a$ on no presumption concerning any other of the propositions in $\mathfrak{A}$ (beyond 0, 1) to the consideration of $a$ on the presumption that $h$ is true, has its exact algebraic counterpart in the passage from $a$ (which may be replaced by $a/1$) to $a/h = a/(\sim h)$. *Henceforth it is in this precise algebraic sense of a remainder class that the eventuality symbol* $a/h$ *shall be employed*, its intuitive counterpart serving solely as a guide to the formulation of the axioms governing the $\prec$ sign, as well as for the interpretation and application of the theorems developed in the formal mathematical structure which we are about to erect.

In closing this section we note, firstly, that the symbol $a/0 = a/(1)$ is useless, since its intuitive content is the assertion of 0: *it shall be implicitly excluded throughout*; and, secondly, that the symbol $\prec$ is regarded as having predicative force, and hence all symbolic constructs into which it enters are asserted, not

contemplated, propositions, and are never denoted by letters nor operated upon with logical constants.[10] Thus it, like $(= \subset)$, has a weaker binding force than $(\sim \cdot \vee)$, so that the parentheses in expressions such as $[(\,)/(\,)] \prec [(\,)/(\,)]$ are omitted. The binding strength of $(/)$ is greater than $(=, \subset, \prec)$ and less than $(\sim \cdot \vee)$.

Finally, it might be interesting to consider eventualities $a/\mathfrak{u}$ in which $\mathfrak{u}$ is not a principal ideal $(\mathfrak{u} \neq (\sim h))$. But a careful examination of this matter has shown us that such an extension is not a useful one at the present stage of the theory.

### 3. The Axioms of Probability

From the standpoint of intuition, the following axioms are the intuitively evident laws of consistency governing all comparisons in probability; in some cases, to be sure, the full understanding of their meaning may offer some difficulty: the only claim is that, their meaning once being apprehended, their truth will be granted.[11] From the standpoint of pure mathematics, they are the formal postulates governing the $\prec$ sign, and are accepted by convention. The axioms form a sufficient basis for what we envision as the legitimate rôle of a theory of intuitive probability. We offer no discussion of consistency, nor have we striven systematically to secure independence (Axiom $S$ has indeed been stated for all integral $n$, whereas if assumed for prime $n$ the general case will follow). In short, the situation is altogether comparable with the case of the laws of intuitive logic.

In each of the following axioms a definite Boolean ring $\mathfrak{A}$ (with units 0, 1) is presupposed in which the remainder classes are formed; and wherever in later discussion a theorem is stated or proved, a definite $\mathfrak{A}$ is again presupposed. But it has seemed unnecessary in general to make explicit mention of this fact in most of the following statements and discussion.

#### The Axioms

V. Axiom of Verified Contingency.
$a/h \prec k/k$.

---

[10] This circumstance shows itself throughout our whole work by the fact that we are never allowed to talk about the probability of a comparison in probability. This may seem to get us unto conflict with Bayes' Rule, which is often regarded as concerning itself with "the probability of a probability." A mere change in wording avoids this difficulty for us, and we regard Bayes' Rule to concern itself only with the probabilities of different values of a physical parameter $\theta$, which, itself, determines the probability (in particular, by the equation $p = \theta$).

[11] A most useful auxiliary in the intuitive comprehension and interpretation of the axioms as well as later theorems and proofs is the logical diagram, in which plane regions represent the propositions, and $(\sim \cdot \vee = \subset)$ are given their set-theoretic rendering. One may go further and let $a/h \prec b/k$ be represented by having the ratio of the area of $ah$ to that of $h \leqq$ that of $bk$ to $k$.

I. AXIOM OF IMPLICATION.
*If* $a/h \succ k/k$, *then* $h \subset a$.

R. AXIOM OF REFLEXIVITY.
$a/h \prec a/h$.

T. AXIOM OF TRANSITIVITY.
*If* $a/h \prec b/k$ *and* $b/k \prec c/l$, *then* $a/h \prec c/l$.

A. AXIOM OF ANTISYMMETRY.
*If* $a/h \prec b/k$, *then* $\sim a/h \succ \sim b/k$.

C. AXIOMS OF COMPOSITION.
*Let* $a_1b_1h_1 \neq 0$ *and* $a_2b_2h_2 \neq 0$.
$C_1$. *If* $a_1/h_1 \prec a_2/h_2$ *and* $b_1/a_1h_1 \prec b_2/a_2h_2$, *then* $a_1b_1/h_1 \prec a_2b_2/h_2$.
$C_2$. *If* $a_1/h_1 \prec b_2/a_2h_2$ *and* $b_1/a_1h_1 \prec a_2/h_2$, *then* $a_1b_1/h_1 \prec a_2b_2/h_2$.

D. AXIOMS OF DECOMPOSITION (QUASI-CONVERSES OF C).
*Let* $a_1b_1h_1 \neq 0$, $a_2b_2h_2 \neq 0$, *and* $a_1b_1/h_1 \prec a_2b_2/h_2$. *Then if either of the eventualities*
(i) $a_1/h_1$, $b_1/a_1h_1$, *has the supraprobable relation* ($\succ$) *with either of*
(ii) $a_2h_2$, $b_2/a_2h_2$, *it will follow that the remaining eventuality of* (i) *will have the infraprobable relation* ($\prec$) *with the remaining one of* (ii).

P. AXIOM OF ALTERNATIVE PRESUMPTION.
*If* $a/bh \prec r/s$ *and* $a/\sim bh \prec r/s$, *then* $a/h \prec r/s$

S. AXIOM OF SUBDIVISION.
*For any integer* $n$ *let the propositions* $a_1, \cdots, a_n, b_1, \cdots, b_n$ *be such that* $a_ia_j = b_ib_j = 0$ $(i \neq j)$ $i, j = 1, \cdots, n$; $a = a_1 \vee \cdots \vee a_n \neq 0$; $b = b_1 \vee \cdots \vee b_n \neq 0$; $a_1/a \prec \cdots \prec a_n/a$; $b_1/b \prec \cdots \prec b_n/b$; *then* $a_1/a \prec b_n/b$.

On fixing the attention upon a given Boolean ring $\mathfrak{A}$ of concrete propositions, in which the $\prec$ relation is supposed to have been introduced, it is seen that the formulation focuses attention upon the totality $\mathfrak{a}$ of remainder classes with respect to all the principal ideals of $\mathfrak{A}$ and that Axioms R and T show that $\prec$ defines a partial ordering of the elements of $\mathfrak{a}$.[12] We shall accordingly appropriate the conventional definitions and theorems of the theory of partially ordered sets.[13] In particular, we shall write in either of the two forms $a/h \prec$

[12] While it is shown at once that if $h \subset k$ (material implication) then $(h) \subset (k)$ (set inclusion), and conversely, we may state at once that there is no such simple relation between the $\prec$ and $\subset$ when they occur between remainder classes. Examples are most easily given after the subject of numerical probability has been developed.

[13] For a full account of this theory, see H. M. MacNeille, *Partially Ordered Sets*, Trans. Amer. Math. Soc., Vol. 42 (1937), pp. 416-460.

In most treatments of partially ordered sets, the greater part of the developments follow from the "Lattice Assumption," which in the present connection would take the following form: If $a/h \in \mathfrak{a}$ and $b/k \in \mathfrak{a}$ then $\mathfrak{a}$ contains $c/l$ having the properties
(i) $c/l \prec a/h$, $c/l \prec b/k$ and

$b/k$, $b/k > a/h$, the abbreviation for the combined assertion of $a/h \prec b/k$ and denial of $a/h \succ b/k$, and we shall use the symbol $\approx$ for *equiprobability*: $a/h \approx b/k$ as the abbreviation for the joint assertion of the two following: $a/h \prec b/k$, $a/h \succ b/k$. (Contrast $\approx$ with $=$: $a/h = b/k$ signifying identity of remainder classes).

A process constantly to be employed in all later work is the application of a factor in the presumption to the contingency; for it is evident in view of the conventions of §2 that, for example $a/hk = ah/hk = ahk/hk$.

It will be remarked that Axiom I is in sharp contrast with the familiar fact that an eventuality which is not regarded as certain (as $a/h$ when $h \not\subset a$) may nevertheless have the same *numerical* probability (unity) as one held as certain (as $k/k$). Our contention is that the truth of Axiom I resides in the very intuitions of comparison in probability and certainty, of which the numerical probability is but an unfaithful representation.

A final remark: If the contingency $a$ is a determinate arithmetical statement, such as "the 100th decimal in $2^{\pi}$ is 2" our theory asserts that the eventuality $a/1$ either $\approx 0/1$ or $1/1$ (cf. Theorem 1 in §4). This is because the truth or falsity of $a$ is determined by the Boolean ring $\mathfrak{A}$ within which all the propositions here envisaged find themselves: in this ring $a$ may be computed by the formal laws of the algebra and found either to $= 0$ or $1$. We hold that this is as it should be and that if the ideas of intuitive probability are to be applied non-trivially to arithmetic, the contingency must be stated in a logically undetermined form which reveals something of the "random process" by which the definition of the integer was settled upon.

## 4. Theorems on Comparison

THEOREM 1. *For all* $a \neq 0$, $a/a \approx 1/1$ *and* $0/a \approx 0/1$ *are true.*
*If* $ah \neq 0$ *and* $ah \neq h$, *then* $0/1 < a/h < 1/1$.

It follows from V that $a/a \prec 1/1$ and $1/1 \prec a/a$, hence $a/a \approx 1/1$. These with A yield $\sim a/a \succ \sim 1/1$ and $\sim 1/1 \succ \sim a/a$; but $\sim a/a = a \sim a/a = 0/a$ and $\sim 1/1 = 0/1$; whence the relation $0/a \approx 0/1$.

From V it follows that $a/h \prec 1/1$; if $a/h \succ 1/1$, I would lead to $ah = h$, contrary to hypothesis; hence $a/h < 1/1$. Now if $\sim ah = h$ it would follow that $0 = a(\sim ah) = ah$, contrary to hypothesis; hence we may replace $a$ by $\sim a$ in the previous discussion and so obtain $\sim a/h \prec 1/1$ together with the denial of $\sim a/h \succ 1/1$; whereupon A shows that $a/h \succ 0/1$ is true and $a/h \prec 0/1$ false; hence $0/1 < a/h$. Thus our theorem is proved.

If in an assertion involving $\prec$ signs, *every* $\prec$ be replaced by $\approx$ a new statement is obtained which shall be called *the sharpening* of the original. If on the other hand *at least one* $\prec$ is replaced by $<$ the new statement shall be called *a strength-*

---

(ii) if $d/m \in \mathfrak{A}$ such that $d/m \prec a/h$ and $d/m \prec b/k$ then $d/m \prec c/l$.
The remaining Lattice Assumption is the $(\prec \rightarrow \succ)$ dual of the above. It is perhaps not unworthy of note that these ideas play no rôle in the present theory of partially ordered sets.

*ening* of the original statement. If a deductive statement (axiom, theorem, corollary, etc.) has a $\prec$ statement for hypothesis and a $\prec$ statement for conclusion, the new deductive statement obtained by replacing both hypothesis and conclusion by their sharpening shall be called the sharpening of the original deductive statement, while the deductive statement obtained by replacing both hypothesis and the conclusion by one of their respective strengthenings shall be called a strengthening of the original deductive statement.

One of our tasks is the establishment of the truth of the sharpening and of every strengthening of every one of the axioms which involves the $\prec$ in both hypothesis and conclusion, to wit, of all except V, I and R. This is conveniently accomplished in two stages announced in Theorems 2 and 11.

THEOREM 2. *The sharpening and every strengthening of Axioms* T, A, C, D *are valid.*

The proofs for T and A are obvious and shall be omitted.

Turning to $C_1$, let the new hypothesis be $a_1b_1h_1 \neq 0$ $a_2b_2h_2 \neq 0$, and $a_1/h_1 \approx a_2/h_2$, $b_1/a_1h_1 \approx b_2/a_2h_2$. We are to prove $a_1b_1/h_1 \approx a_2b_2/h_2$. Now the hypothesis leads to the original hypothesis of $C_1$, and also to this latter in which the subscripts 1 and 2 have been interchanged; thus $C_1$ leads to the two conclusions $a_1b_1/h_1 \prec a_2b_2/h_2$, $a_1b_1/h_1 \succ a_2b_2/h_2$, which is equivalent to the conclusion sought.

As for the strengthening of $C_1$, let the new hypothesis be $a_1b_1h_1 \neq 0$, $a_2b_2h_2 \neq 0$, and $a_1/h_1 < a_2/h_2$, $b_1/a_1h_1 \prec b_2/a_2h_2$; we are to prove $b_1/a_1h_1 < a_2/h_2$. In view of $C_1$ all that is required is to show the falsity of $a_1b_1/h_1 \succ a_2b_2/h_2$. Let this relation be assumed true and apply D (with subscripts interchanged) to the two relations $b_1/a_1h_1 \prec b_2/a_2h_2$, $a_1b_1/h_1 \succ a_2b_2/h_2$, thus deriving the following contradiction of hypothesis: $a_2/h_2 \prec a_1/h_1$.

The remaining possibility (not reducible to the above) is the introduction into the hypothesis of $C_1$ of $b_1/a_1h_1 < b_2/a_2h_2$. The treatment of this case being quite similar to the other is left to the reader.

The discussion of $C_2$ is omitted, being altogether similar to the case of $C_1$. And as for D, we shall treat only the following case, leaving the rest to the reader:

Let $a_1b_1h_1 \neq 0$, $a_2b_2h_2 \neq 0$, and $a_1b_1/h_1 < a_2b_2/h_2$, $a_1/h_1 \succ b_2/a_2h_2$. We are to prove that $b_1/a_1h_1 < a_2/h_2$. It is required only to show the falsity of $b_1/a_1h_1 \succ a_2/h_2$. For this purpose we have but to apply $C_2$ (subscripts interchanged) to the relations $a_2/h_2 \prec b_1/a_1h_1$, $b_2/a_2h_2 \prec a_1/h_1$, whereupon there will follow the contradiction of hypothesis: $a_2b_2/h_2 \prec a_1b_1/h_1$.

In the succeeding theorems, when it is desired to indicate that sharpening and strengthenings are valid, a statement to this effect is appended to the statement of the theorem. But in those cases where the proof of the sharpened or strengthened theorem is furnished by simply paraphrasing that of the original theorem, with the replacement of axioms, theorems, etc., employed by their sharpening or strengthenings, the proof is omitted. And wherever in the future reference is made to an axiom or theorem, either the original or its sharpening or strengthen-

ing shall be meant: the reader will have no difficulty in seeing which one is needed.

THEOREM 3. *If* $ah \subset bh$ *then* $a/h \prec b/h$. *If in addition,* $bh \not\subset ah$, *then* $a/h < b/h$.

From R and V it follows that $b/h \prec b/h$, $a/bh \prec b/bh$. It is now possible to apply $C_1$ in which we put $h_1 = h_2 = h$, $a_1 = a_2 = b$, $b_1 = a$, $b_2 = b$, and so to derive the following relation, which furnishes the first conclusion desired, $ab/h \prec bb/h$. Here $a_1b_1h_1 = abh = ah$ and $a_2b_2h_2 = bh$, so that if $a_2b_2h_2 = 0$ we should have $ah = abh = 0$, in which case Theorem 1 would furnish the desired conclusion.

If, further, it is assumed that $bh \not\subset ah$, whereas $a/h \not\prec b/h$, i.e., $a/h \succ b/h$, we could apply D to the following relations $b/h \prec a/h$, $b/h \succ b/h$, where we put $h_1 = h_2 = h$, $a_1 = a_2 = b$, $b_1 = b$, $b_2 = a$, and so either $a_1b_1h_1(=bh)$ and $a_2b_2h_2(=abh=ah)$ are not both $\neq 0$ or else the relation $b/bh \prec a/bh$ is obtained. But we have $b/bh = bh/bh \approx 1/1$, hence this would give by Theorem 1 that $bh \subset a$, so that $bh \subset ah$, in contradiction with the hypothesis. If $ah = 0$ then in view of the relation we are trying to disprove, we should have $b/h \prec a/h = 0/h \approx 0/1$ and thus by Theorem 1, $bh = 0 \subset ah$. Finally, if $bh = 0$, the same contradiction with the hypothesis $bh \not\subset ah$ would obtain. Thus all parts of the theorem are proved.

The second part of this theorem may be regarded as a kind of strengthening of the first part. The sharpening in the sense that $ah = bh$ implies $ah/h \approx bh/h$ is a trivial consequence of R. Similar remarks apply to the following theorem.

THEOREM 4. *If* $a \subset k \subset h$, *then* $a/h \prec a/k$. *If in addition* $h \not\subset k$ *and* $a \neq 0$, *then* $a/h < a/k$.

From V and R it follows that $k/h \prec k/k$, $a/hk \prec a/hk$. It is possible to apply $C_1$ after setting $h_1 = h$, $h_2 = a_1 = a_2 = k$, $b_1 = b_2 = a$, from which the desired conclusion follows. Here $a_1b_1h_1 = ahk = a$, $a_2b_2h_2 = ak = a$, and if $a = 0$ the conclusion would follow from Theorem 1.

If, furthermore, $h \not\subset k$ and $a \neq 0$, we could apply D to the supposed relations $a/k \prec a/h$, $a/k \succ a/k$, where we set $h_1 = k$, $h_2 = h$, $a_1 = a_2 = k$, $b_1 = b_2 = a$, from which it would follow,—providing $a_1b_1h_1(=ak=a)$ and $a_2b_2h_2(=akh=a)$ are neither $= 0$, a case excluded by hypothesis,—that $1/1 \approx k/k \prec k/h$; thus $h \subset k$, and we have a contradiction.

THEOREM 5. *If* (i) $a_1b_1h_1 = a_2b_2h_2 = 0$, *and if* (ii) $a_1/h_1 \prec a_2/h_2$, (iii) $b_1/h_1 \prec b_2/h_2$, *then* $a_1 \vee b_1/h_1 \prec a_2 \vee b_2/h_2$.

*The sharpening and all strengthenings are valid.*

On applying A to (ii) one obtains

$$\sim a_1/h_1 \succ \sim a_2/h_2 . \tag{1}$$

From (i) it follows that $\sim a_1 \vee \sim b_1 \vee \sim h_1 = 1$, whence, multiplying through by $b_1h_1$: $\sim a_1b_1h_1 = b_1h_1$; thus we have $\sim a_1b_1/h_1 = \sim a_1b_1h_1/h_1 = b_1h_1/h_1 = b_1/h_1$, and similarly for $b_2/h_2$. Hence (iii) may be written as

$$\sim a_1b_1/h_1 \prec \sim a_2b_2/h_2 . \tag{2}$$

It may be assumed that $\sim a_1 b_1 h_1 \neq 0$ and $\sim a_2 b_2 h_2 \neq 0$. For if $\sim a_1 b_1 h_1 = 0$, (i) would give us $b_1 h_1 = \sim a_1 b_1 h_1 \vee a_1 b_1 h = 0$, and the conclusion would follow directly from the following relations made evident by (ii) and Theorem 3: $a_1 \vee b_1/h_1 = a_1 h_1 \vee b_1 h_1/h_1 = a_1 h_1/h_1 = a_1/h_1$ ; $a_1/h_1 \prec a_2/h_2 \prec a_2 \vee b_2/h_2$. While if $\sim a_2 b_2 h_2 = 0$, we could derive $b_2 h_2 = 0$ as before, whereupon (iii) would yield with the aid of Theorem 1 the already excluded case $\sim a_1 b_1 h_1 = 0$. We are, therefore, entitled to apply D to (1) and (2) ($\sim a_1$ , $\sim a_2$ playing the rôle of "$a_1$ , $a_2$") and so obtain

$$b_1/\sim a_1 h_1 \prec b_2/\sim a_2 h_2 . \tag{3}$$

This with A gives the relation $\sim b_1/\sim a_1 h_1 \succ \sim b_2/\sim a_2 h_2$. If $\sim a_1 \sim b_1 h_1 \neq 0$ and $\sim a_2 \sim b_2 h_2 \neq 0$, $C_1$ may be applied to this relation together with (1), ($\sim a_2$ , $\sim b_2$ , $h_2$ , $\sim a_1$ , $\sim b_1$ , $h_1$ playing the rôle of "$a_1$ , $b_1$ , $h_1$ , $a_2$ , $b_2$ , $h_2$"), whereupon $\sim a_1 \sim b_1/h_1 \succ \sim a_2 \sim b_2/h_2$ ,which, on application of A, yields the conclusion of our theorem.

If on the other hand $\sim a_2 \sim b_2 h_2 = 0$, it would follow that $a_2 \vee b_2 \vee \sim h_2 = 1$, whence $h_2 = h_2(a_2 \vee b_2 \vee \sim h_2) = (a_2 \vee b_2) h_2$ , and thus $a_2 \vee b_2/h_2 = (a_2 \vee b_2) h_2/h_2 = h_2/h_2$ and our conclusion would be a consequence of V. Finally, if $\sim a_1 \sim b_1 h_1 = 0$ we should have (as $\sim a_1 b_1 h_1 \neq 0$ is assumed) $\sim a_1 h_1 = \sim a_1 \sim b_1 h_1 \vee \sim a_1 b_1 h_1 = \sim a_1 b_1 h_1$ , and thus by Theorem 1, $b_1/\sim a_1 h_1 = \sim a_1 b_1 h_1/a_1 h_1 \approx 1/1$, which in conjuction with (3) gives $b_2/\sim a_2 h_2 \succ 1/1$ whence, again by Theorem 1, $\sim a_2 b_2 h_2 = \sim a_2 h_2$ and multiplication of this by $\sim b_2$ gives $\sim a_2 \sim b_2 h = 0$, the case previously disposed of. Thus our proof is complete.

THEOREM 6. *If $a_{1i} a_{1j} h_1 = a_{2i} a_{2j} h_2 = 0$ for all $i \neq j$, and if $a_{1i}/h_1 \prec a_{2i}/h_2$ for all $i = 1, 2, \cdots, n$, then $a_{11} \vee a_{12} \vee \cdots \vee a_{1n}/h_1 \prec a_{21} \vee a_{22} \vee \cdots \vee a_{2n}/h_2$.*

*The sharpening and all strengthenings are valid.*

This follows from Theorem 5 by mathematical induction.

THEOREM 7. *If* (i) $b_1 \subset a_1$ , $b_2 \subset a_2$ *and if* (ii) $a_1/h_1 \prec a_2/h_2$ , (iii) $b_1/h_1 \succ b_2/h_2$ *then* $a_1 \sim b_1/h_1 \prec a_2 \sim b_2/h_2$ .

*The sharpening and all strengthenings are valid.*

In virtue of (i), $a_1 b_1 = b_1$ and $a_2 b_2 = b_2$ , so that (iii) becomes $a_1 b_1/h_1 \succ a_2 b_2/h_2$. Suppose firstly that $a_1 b_1 h_1 \neq 0$, $a_2 b_2 h_2 \neq 0$. Then the application of D (with subscripts interchanged) to this relation together with (ii) yields $b_1/a_1 h_1 \succ b_2/a_2 h_2$ , which on application of A furnishes the relation

$$\sim b_1/a_1 h_1 \prec \sim b_2/a_2 h_2 . \tag{4}$$

It is now permitted to assume that $a_1 \sim b_1 h_1 \neq 0$, $a_2 \sim b_2 h_2 \neq 0$. For if $a_1 \sim b_1 h_1 = 0$, then $a_1 \sim b_1/h_1 = a_1 \sim b_1 h_1/h_1 = 0/h_1 \approx 0/1$, and the conclusion of our theorem would be a consequence of Theorem 1. And if $a_2 \sim b_2 h_2 = 0$, relation (4) would give $\sim b_1/a_1 h_1 \prec a_2 \sim b_2 h_2/h_2 = 0/h_2 \approx 0/1$, whence Theorem 1 leads to the case $a_1 \sim b_1 h_1 = 0$ just disposed of. We may, then, apply $C_1$ to (4) in conjunction with (ii) (with $\sim b_1$ , $\sim b_2$ replacing "$b_1$ , $b_2$"), and so obtain the conclusion of our theorem.

Turning lastly to the excluded cases, suppose that $a_2 b_2 h_2 = 0$. This gives in

view of (i) $b_2h_2 = 0$, so that $\sim b_2 \vee \sim h_2 = 1$, $\sim b_2h_2 = h_2$ ; thus $a_2 \sim b_2/h_2 = a_2(\sim b_2h_2)/h_2 = a_2h_2/h_2 = a_2/h_2$ . Now $a_1 \sim b_1 \subset a_1$ , and we have by Theorem 3 $a_1 \sim b_1/h_1 \prec a_1/h_1$ . The conclusion to our theorem then follows from (ii). Finally, if $a_1b_1h_1 = 0$, (i) gives that $b_1h_1 = 0$, and thus $b_1/h_1 = b_1h_1/h_1 = 0/h_1 \approx 0/1$; but from (iii) it follows that $b_2/h_2 \prec 0/1$ whence $b_2h_2 = 0$,—the case just disposed of.

THEOREM 8. *Let $a \subset k$, $b \subset k$, $k \subset h$. Then either of the following relations is a necessary and sufficient condition for the truth of the other*:

(i) $a/h \prec b/h$; (ii) $a/k \prec b/k$.

*The sharpening and strengthening are valid.*

Assume (i). Since $ka = a$ and $kb = b$, we have on account of (i) and R the relations $ka/h \prec kb/h$, $k/h \succ k/h$. Now $kah = a$ and $kbh = b$, and the supposition that either or both of these $= 0$ leads at once to (ii) (Theorem 1). We may accordingly apply D to the above pair of relations (with $k$, $a$, $h$, $k$, $b$, $h$, replacing $a_1$ , $b_1$ , $h_1$ , $a_2$ , $b_2$ , $h_2$ , respectively) and so obtain $a/kh \prec b/kh$, which, in view of $kh = k$, is the desired conclusion (ii).

Assume (ii). We may apply $C_1$ to the two relations $k/h \prec k/h$, $a/kh \prec b/kh$, (with $k$, $a$, $h$, $k$, $b$, $h$, replacing $a_1$ , $b_1$ , $h_1$ , $a_2$ , $b_2$ , $h_2$ , respectively) and so obtain a relation which, since $ka = a$ and $kb = b$, is identical with (i). Here again $kah = a$ and $kbh = b$, and if either $= 0$ the conclusion is trivial.

THEOREM 9. *Let $ah \neq 0$ and $bh \neq 0$. Then from*

(i) $$a/h \prec a/bh$$

*it will follow that*

(ii) $$b/h \prec b/ah.$$

*The theorem remains valid if the $\prec$ be replaced by $\succ$ in both* (i) *and* (ii).

*The sharpening and strengthening are valid.*

We may write (i) and a relation evident from R as the pair $ab/h \prec ab/h$, $a/bh \succ a/h$, and may then apply D (with $b$, $a$, $h$, $a$, $b$, $h$, replacing $a_1$ , $b_1$ , $h_1$ , $a_2$ , $b_2$ , $h_2$ , respectively), whereupon (ii) is obtained. That this application of D may be justified, we observe that $a_1b_1h_1 = abh = a_2b_2h_2$ ; suppose that $abh = 0$; then $b/ah = abh/ah = 0/ah \approx 0/1$, and (ii) becomes a trivial consequence of Theorem 1.

The proof of the remainder of the theorem follows precisely similar lines and may be omitted.

THEOREM 10. *Let $kh \neq 0$ and $\sim kh \neq 0$. Then either of the following relations is a necessary and sufficient condition for the truth of the other*:

$$\text{(i) } a/h \prec a/kh; \qquad \text{(ii) } a/h \succ a/\sim kh.$$

*The sharpening and strengthening are valid.*

It is only necessary for us to show that (i) implies (ii), in view of A. If $ah = 0$, then $a/\sim kh = ah/\sim kh = 0/\sim kh \approx 0/1$, and (ii) would be a consequence of Theorem 1. If $ah \neq 0$, Theorem 9 may be applied to (i) (with $b = k$) and thus a

relation obtained which under the action of A yields $\sim k/h \succ \sim k/ah$. Again applying Theorem 9 (with $\sim k$, $a$, $\succ$, replacing $a$, $b$, $\prec$), we derive (ii).

THEOREM 11. *The sharpening and all strengthenings of Axioms* P *and* S *are valid.*

For the sharpening of P, the hypothesis $a/bh \approx r/s$, $a/\sim bh \approx r/s$ on the one hand contains the hypothesis of P and hence its conclusion, and on the other hand contains the relations $a/bh \succ r/s$, $a/\sim bh \succ r/s$, whence it follows by A that $\sim a/bh \prec \sim r/s$, $\sim a/\sim bh \prec \sim r/s$, whence again by P, $\sim a/b \prec \sim r/s$, and again by A, $a/b \succ r/s$, which in conjunction with the former conclusion of P gives the sharpened conclusion.

For the strengthening of P, let it be assumed that $a/bh \prec r/s$, $a/\sim bh \prec r/s$. All that is required is to prove the falsity of

$$a/h \succ r/s. \tag{5}$$

Assume this to be true; our hypothesis yields with the aid of T, $a/\sim bh \prec a/h$. Now apply Theorem 10,—the above relation being equivalent to (ii) with $k = b$,—and we have $a/h \prec a/bh$, which in conjunction with (5) gives the contradiction of hypothesis $a/bh \succ r/s$. The other case is automatically settled by the foregoing.

The sharpening of S is established at once on the double application of S to the $\prec$ and the $\succ$ hypotheses contained in the sharpened hypothesis.

For the strengthening of S, suppose the strengthened hypothesis to contain for some fixed $i (1 \leqq i \leqq n-1)$ $a_i/a \prec a_{i+1}/a$, we are to prove $a_1/a \prec b_n/b$. Assume on the contrary that $a_1/a \succ b_n/b$, it follows at once that

$$a_1/a \succ b_1/b, \quad \cdots, \quad a_i/a \succ b_i/b,$$

$$a_{i+1}/a \succ b_{i+1}/b, \quad \cdots, \quad a_n/a \succ b_n/b;$$

whence in view of Theorem 6 (strengthened) the contradiction follows $1/1 \approx a/a \succ b/b \approx 1/1$. The remaining cases are left to the consideration of the reader.

THEOREM 12. *Let* $ah \neq 0$, $\sim ah \neq 0$, $bh \neq 0$, $\sim bh \neq 0$. *Then from*

$$a/bh \prec a/\sim bh \tag{i}$$

*it follows that*

$$b/ah \prec b/\sim ah \tag{ii}$$

$$a/bh \prec a/h. \tag{iii}$$

*The theorem is valid if* $\prec$ *be replaced by* $\succ$ *in every one of* (i), (ii), (iii).

*The sharpening and strengthening are valid.*

We have in view of R that $a/\sim bh \prec a/\sim bh$. Apply P to this in conjunction with (i) (with $a$, $\sim bh$, replacing $r$, $s$); we obtain $a/h \prec a/\sim bh$. Now we apply Theorem 10 (with $k = \sim b$), and so have (iii). The relation (ii) now follows by T.

The following extension of P may be given; we omit the proof which is immediate.

THEOREM 13. *Under the hypothesis that*

$$h_ih_j = 0 : (i \neq j)\ i, j = 1, \cdots, m;$$

$$k_ik_j = 0 : (i \neq j)\ i, j = 1, \cdots, n;$$

$$a/h_i \prec b/k_j\ ;\quad i = 1, \cdots, m;\quad j = 1, \cdots, n.$$

*It follows that*

$$a/h_1 \vee \cdots \vee h_m \prec b/k_1 \vee \cdots \vee k_n\ .$$

*The sharpening and strengthenings are valid.*

## 5. Numerical Probability

The object of this section is to lay the foundations of the theory of numerical probability, or mathematical theory of probability in the conventional sense, upon the basis of the theory expounded in the preceding sections so that the classical developments can be made to follow from the present work. In addition to the axioms of §3 it is necessary to make a further assumption, the precise philosophical character of which we shall not undertake to examine, concerning the existence of what we shall call $n$-scales.

DEFINITION 1. *By an $n$-scale shall be meant any set of $n$ propositions $(u_1, \cdots, u_n)$ for which the following is posited*:

1. $u = u_1 \vee \cdots \vee u_n \neq 0$.
2. $u_iu_j = 0$ $(i \neq j)$ for all $i, j = 1, \cdots, n$.
3. $u_i/u \approx u_j/u$ for all $i, j = 1, \cdots, n$.

ASSUMPTION. *If $n$ is any given positive integer, at least one $n$-scale may be regarded as existing.*

THEOREM 14. *If $(u_1, \cdots, u_n)$ is an $n$-scale and $(v_1, \cdots, v_m)$ an $m$-scale, and if $\rho$ and $\sigma$ are integers with $0 \leqq \sigma \leqq n, 0 \leqq \rho \leqq m$, then*

$$u_{i_1} \vee \cdots \vee u_{i_\sigma}/u \ <, \approx, > \ v_{j_1} \vee \cdots \vee v_{j_\rho}/v$$

*according as $\sigma/n <, =, > \rho/m$.*

Here it is understood that $i_1, \cdots, i_\sigma$ are each between 1 and $n$, and that no two are equal; similarly for $j_1, \cdots, j_\rho$ and $m$. And, finally, when $\sigma = 0$ the first member is replaced by $0/u$, and when $\rho = 0$, the second by $0/v$; the case where either or both of these are true is trivial, and shall be left out of consideration in our proof.

Since by hypothesis

$$u_{i_1}/u \approx u_1/u, \cdots, u_{i_\sigma}/u \approx u_\sigma/u,$$

$$v_{j_1}/v \approx v_1/v, \cdots, v_{j_\rho}/v \approx v_\rho/v,$$

we have by Theorem 6 (sharpened) that

$$u_{i_1} \vee \cdots \vee u_{i_\sigma}/u \approx u_1 \vee \cdots \vee u_\sigma/u,$$

$$v_{j_1} \vee \cdots \vee v_{j_\rho}/v \approx v_1 \vee \cdots \vee v_\rho/v,$$

and thus it suffices to prove that

$$u_1 \vee \cdots \vee u_\sigma/u <, \approx, > v_1 \vee \cdots \vee v_\rho/v$$

according as $\sigma/n <, \approx, > \rho/m$.

In virtue of our underlying assumption, we are permitted posit the existence of an $mn$-scale $(w_1, \cdots, w_{mn})$. Let $U_i$ $(i = 1, \cdots, n)$ and $V_j$ $(j = 1, \cdots, m)$ be defined as follows

$$U_i = w_{m(i-1)+1} \vee w_{m(i-1)+2} \vee \cdots \vee w_{mi},$$

$$V_j = w_{n(j-1)+1} \vee w_{n(j-1)+2} \vee \cdots \vee w_{nj}.$$

Clearly we have

$$w = w_1 \vee \cdots \vee w_{mn} = U_1 \vee \cdots \vee U_n = V_1 \vee \cdots \vee V_m;$$

$$U_iU_j = 0, \qquad V_iV_j = 0, \qquad (i \neq j);$$

$$U_i/w \approx U_j/w, \qquad V_i/w \approx V_j/w;$$

the last, again by Theorem 6. Thus we may apply Axiom S to $u_1, \cdots, u_n$ and $U_1, \cdots, U_n$, and so we obtain $u_i/u \approx U_j/w$, $(i, j = 1, \cdots, n$. Similarly, for $v_1, \cdots, v_m$ and $V_1, \cdots, V_m$, we have $v_i/v \approx V_j/w$, $i, j = 1, \cdots, m$. Applying Theorem 6 again, we derive

$$u_1 \vee \cdots \vee u_\sigma/u \approx U_1 \vee \cdots \vee U_\sigma/w$$

$$v_1 \vee \cdots \vee v_\rho/v \approx V_1 \vee \cdots \vee V_\rho/w.$$

Now

$$U_1 \vee \cdots \vee U_\sigma = w_1 \vee \cdots \vee w_{m\sigma},$$

$$V_1 \vee \cdots \vee V_\rho = w_1 \vee \cdots \vee w_{n\rho},$$

and thus by Theorem 3

$$u_1 \vee \cdots \vee u_\rho/u <, \approx, > v_1 \vee \cdots \vee v_\sigma/v$$

according as

$$w_1 \vee \cdots \vee w_{m\sigma} \subset \text{and } \not\supset, = , \supset \text{and } \not\subset w_1 \vee \cdots \vee w_{n\rho},$$

that is, as $m\sigma <, =, > n\rho$, from which the theorem follows.

Fixing the attention upon a given eventuality $a/h$, let $n$ be an arbitrary positive integer, and $(u_1, \cdots, u_n)$ a certain $n$-scale. The relation $u_1 \vee \cdots \vee u_t/u \prec a/h$ will be true for at least one value of $t$ $(0 \leqq t \leqq n)$, if we agree to include the case $0/u \prec a/h$ for $t = 0$. Furthermore, the maximum value of $t$ for which this relation holds is independent of the particular $n$-scale $(u_1, \cdots, u_n)$ chosen, as an obvious application of Theorem 14 will show: This value of $t$ is thus determined when $n$ is given (for fixed $a/h$) and shall be denoted by $t(n)$.

Similarly, the minimum value of $T$ for which $a/h \prec u_1 \vee \cdots \vee u_T/u$ is

true is determined when $n$ is given (for fixed $a/h$), and shall be denoted by $T(n)$. Evidently we have $0 \leqq t(n) \leqq T(n) \leqq n$.

THEOREM 15. *The following limits always exist*:

$$p_*(a, h) = \lim_{n\to\infty} \frac{t(n)}{n}; \qquad p^*(a, h) = \lim_{n\to\infty} \frac{T(n)}{n};$$

*and they satisfy the relation*

$$0 \leqq p_*(a, h) \leqq p^*(a, h) \leqq 1.$$

Let $n$, $m$ be any positive integers, and $(u_1, \cdots, u_n)$ and $(v_1, \cdots, v_m)$ any $n$-scale and $m$-scale. By definition of $t(n)$, we can *assert*

$$u_1 \vee \cdots \vee u_{t(n)}/u \prec a/h$$

and *deny*

$$u_1 \vee \cdots \vee u_{t(n)+1}/u \prec a/h.$$

Hence we must have

$$\frac{t(m)}{m} \leqq \frac{t(n)+1}{n};$$

for otherwise, by Theorem 14, we should have

$$u_1 \vee \cdots \vee u_{t(n)+1}/u \prec v_1 \vee \cdots \vee v_{t(m)}/v,$$

whereupon a contradiction is produced in view of the known relation

$$v_1 \vee \cdots \vee v_{t(m)}/v \prec a/h.$$

Similarly we prove that

$$\frac{t(n)}{n} \leqq \frac{t(m)+1}{m}.$$

On combining these we have

$$-\frac{1}{m} \leqq \frac{t(m)}{m} - \frac{t(n)}{n} \leqq \frac{1}{n},$$

and thus $t(m)/m - t(n)/n \to 0$ as $n, m \to \infty$, and the existence of $p_*(a, h)$ is established. Similarly for $p^*(a, h)$. The relation $0 \leqq p_* \leqq p^* \leqq 1$ is an immediate consequence of $0 \leqq t(n) \leqq T(n) \leqq n$.

THEOREM 16. *The relation $a_1/h_1 \prec a_2/h_2$ implies both $p_*(a_1, h_1) \leqq p_*(a_2, h_2)$ and $p^*(a_1, h_1) \leqq p^*(a_2, h_2)$. And $p^*(a_1, h_1) < p_*(a_2, h_2)$ implies the relation $a_1/h_1 \prec a_2/h_2$.*

The first part of the theorem is obvious. To prove the second, we observe that since

$$\lim_{n\to\infty} \frac{T_1(n)}{n} = p^*(a_1, h_1) < p_*(a_2, h_2) = \lim_{n\to\infty} \frac{t_2(n)}{n},$$

we may take $n$ such that $T_1(n)/n < t_2(n)/n$, whereupon

$$a_1/h_1 \prec u_1 \vee \cdots \vee u_{T_1(n)}/u < u_1 \vee \cdots \vee u_{t_2(n)}/u \prec a_2/h_2 ,$$

in virtue of Theorem 14.

THEOREM 17. $p_*(a, h) + p^*(\sim a, h) = 1$.

Let $(u_1 , \cdots , u_n)$ be an $n$-scale, and observe that for any $s$ $(0 \leqq s \leqq n)$,

$$\sim(u_1 \vee \cdots \vee u_s)/u = u_{s+1} \vee \cdots \vee u_n/u \approx u_1 \vee \cdots \vee u_{n-s}/u.$$

Thus with the aid of Axiom A and the relation

$$u_1 \vee \cdots \vee u_{t(n)}/u \prec a/h \prec u_1 \vee \cdots \vee u_{T(n)}/u$$

we derive the formula

$$u_1 \vee \cdots \vee u_{n-T(n)}/u \prec \sim a/h \prec u_1 \vee \cdots \vee u_{n-t(n)}/u.$$

Comparing this with the relation defining $t'(n)$, $T'(n)$ (for $\sim a/h$):

$$u_1 \vee \cdots \vee u_{t'(n)}/u \prec \sim a/h \prec u_1 \vee \cdots \vee u_{T'(n)}/u,$$

we have $t'(n) \geqq n - T(n)$ and $T'(n) \leqq n - t(n)$. Dividing by $n$ and letting $n \to \infty$ we have

$$p_*(\sim a, h) + p^*(a, h) \geqq 1, \qquad p^*(\sim a, h) + p_*(a, h) \leqq 1.$$

Now the corresponding relations with $a$ and $\sim a$ interchanged are evidently valid. Thus $1 \leqq p_*(a, h) + p^*(\sim a, h) \leqq 1$, and our theorem is proved.

THEOREM 18. (i) $p^*(a_1 \vee a_2 , h) \leqq p^*(a_1 , h) + p^*(a_2 , h)$. (ii) *If* $a_1 a_2 h = 0$, *then* $p_*(a_1 \vee a_2 , h) \geqq p_*(a_1 , h) + p_*(a_2 , h)$.

Let $t(n)$, $T(n)$ correspond to $a_1 \vee a_2/h$, and $t_1(n)$, $T_1(n)$ correspond to $a_1/h$, and, finally, $t_2(n)$, $T_2(n)$ correspond to $a_2/h$.

If for infinitely many values of $n$, $T_1(n) + T_2(n) \geqq n$, dividing by $n$ and allowing $n \to \infty$ through such values would show that $p^*(a_1 , h) + p^*(a_2 , h) = 1$, and conclusion (i) becomes obvious. On the contrary assumption, for all sufficiently large $n$, $T_1(n) + T_2(n) < n$, and we should have

$$a_1/h \prec u_1 \vee \cdots \vee u_{T_1(n)}/u,$$

$$a_2/h \prec u_1 \vee \cdots \vee u_{T_2(n)}/u \approx u_{T_1(n)+1} \vee \cdots \vee u_{T_1(n)+T_2(n)}/u,$$

and hence, by Theorem 5

$$a_1 \vee a_2/h \prec u_1 \vee \cdots \vee u_{T_1(n)+T_2(n)}/u.$$

(In this application of Theorem 5, the requirement that $a_1 a_2 h = 0$ is dispensed with, inasmuch as we replace $a_2$ by $a_2 \sim a_1$, and can then replace $a_2/h$ in the above relation by $a_2 \sim a_1/h$ which, by Theorem 3, $\prec a_2/h$.) Thus we obtain $T(n) \leqq T_1(n) + T_2(n)$, whence the conclusion (i).

Turning now to (ii), suppose that for some $n$, $t_1(n) + t_2(n) \geqq n$. We should have

$$u_1 \vee \cdots \vee u_{t_1(n)}/u \prec a_1/h,$$

$$u_{t_1(n)+1} \vee \cdots \vee u_n/u \prec u_1 \vee \cdots \vee u_{t_2(n)}/u \prec a_2/h,$$

and thus by Theorem 5, $u/u \prec a_1 \vee a_2/h$, hence by Theorems 1 and 16, $p_*(a_1 \vee a_2, h) = 1$, and (ii) is established. On the contrary assumption, $t_1(n) + t_2(n) < n$, and we have

$$u_{t_1(n)+1} \vee \cdots \vee u_{t_1(n)+t_2(n)}/u \approx u_1 \vee \cdots \vee u_{t_2(n)}/u \prec a_2/h$$

which, with the other relation, again by Theorem 5, yields

$$u_1 \vee \cdots \vee u_{t_1(n)+t_2(n)}/u \prec a_1 \vee a_2/h.$$

Hence $t_1(n) + t_2(n) \leqq t(n)$, and the conclusion (ii) follows.

THEOREM 19. *If $a$, $b$, $c$ are any propositions for which $a \subset b \subset c$ ($b$, $c \neq 0$), then*

(i) $$p_*(a, b)p_*(b, c) \leqq p_*(a, c) \leqq p_*(a, b)p^*(b, c)$$

(ii) $$p^*(a, b)p_*(b, c) \leqq p^*(a, c) \leqq p^*(a, b)p^*(b, c).$$

We shall assume in the proof that $a \neq 0$, the contrary case being trivial.

Let $t_1(n)$, $T_1(n)$ correspond to $a/c$, let $t_2(n)$, $T_2(n)$ correspond to $b/c$, and let $t_3(n)$, $T_3(n)$ correspond to $a/b$.

From $a \subset b$ it follows by Theorem 3 that $a/c \prec b/c$, whence by Theorem 16 that

(1) $$p_*(a, c) \leqq p_*(b, c), \qquad p^*(a, c) \leqq p^*(b, c).$$

We shall first establish the inequality

(2) $$p_*(a, b)p_*(b, c) \leqq p_*(a, c).$$

We may evidently assume that $t_2(n) \to \infty$ as $n \to \infty$, as otherwise the conclusion $p_*(b, c) = 0$ could be drawn from Theorem 15, and (2) would be trivial. Clearly, for $t_2(n) \neq 0$, $(u_1, \cdots, u_{t_2(n)})$ forms a $t_2(n)$-scale: For if $u_1 \vee \cdots \vee u_{t_2(n)} = 0$ it would follow that $u_1 = 0$; hence that $u_i/u \approx u_1/u \approx 0/u$; whence by Theorem 6 that $u/u \approx 0/u$, contrary to Theorem 1. And the equiprobability of all expressions

$$u_i/u_1 \vee \cdots \vee u_{t_2(n)}, \qquad 1 \leqq i \leqq t_2(n),$$

is a consequence of that of all the $u_i/u$ (Theorem 8, with $a$, $b$, $k$, $h$ replaced by $u_i$, $u_j$, $u_1 \vee \cdots \vee u_{t_2(n)}$, $u$, respectively). Consequently we may write

$$u_1 \vee \cdots \vee u_{t_3[t_2(n)]}/u_1 \vee \cdots \vee u_{t_2(n)} \prec a/b.$$

This may be combined with

$$u_1 \vee \cdots \vee u_{t_2(n)}/u \prec b/c,$$

the application of Axiom $C_1$ giving, upon setting

$$a_1 , b_1 , h_1 = u_1 \vee \cdots \vee u_{t_2(n)} , u_1 \vee \cdots \vee u_{t_3[t_2(n)]} , u;$$

$$a_2 , b_2 , h_2 = b, a, c,$$

the relation

$$u_1 \vee \cdots \vee u_{t_3[t_2(n)]}/u \prec a/c;$$

from which by definition of $t_1(n)$ we conclude that

$$t_3[t_2(n)] \leqq t_1(n).$$

Since $t_2(n) \to \infty$ with $n$, we have for sufficiently large $n$

$$\frac{t_2(n)}{n} \frac{t_3[t_2(n)]}{t_2(n)} \leqq \frac{t_1(n)}{n},$$

and on letting $n \to \infty$ the inequality (2) is obtained.

In a precisely similar manner, we establish the inequality

$$p^*(a, c) \leqq p^*(a, b)p^*(b, c). \tag{3}$$

In this case, the failure of the assumption $T_2(n) \to \infty$ with $n$ would lead to $p^*(b, c) = 0$ and hence by inequality (1), to $p^*(a, c) = 0$, and thus trivially to (3). The remainder of the proof of (3) follows that of (2).

We turn now to the proof of the inequality

$$p^*(a, b)p_*(b, c) \leqq p^*(a, c). \tag{4}$$

Suppose that for infinitely many values of $n$, $t_2(n) \leqq T_1(n)$: it will follow at once that $p_*(b, c) \leqq p^*(a, c)$, whence (4) results immediately, inasmuch as $p^*(a, b) \leqq 1$. Suppose on the contrary that $T_1(n) < t_2(n)$ for all sufficiently large $n$. We invoke Axiom D in which we set

$$a_1 = b, \quad b_1 = a, \quad h_1 = c, \quad h_2 = u,$$

$$a_2 = u_1 \vee \cdots \vee u_{t_2(n)} , \quad b_2 = u_1 \vee \cdots \vee u_{T_1(n)} .$$

In virtue of $a \subset b \subset c$ and the definition of $t_2(n)$, $T_1(n)$, we have

$$a_1b_1/h_1 \prec a_2b_2/h_2 , \quad a_1/h_1 \succ a_2/h_2 ;$$

furthermore, we have effectively excluded the possibility of either of the equations (since $T_1(n) = 0$ would imply $a/c \approx 0/1$, hence that $a = 0$)

$$a_1b_1h_1 \ (= a) = 0, \quad a_2b_2h_2 \ (= u_1 \vee \cdots \vee u_{T_1(n)}) = 0;$$

and thus the conclusion of Axiom D permits us to write

$$a/b \prec u_1 \vee \cdots \vee u_{T_1(n)}/u_1 \vee \cdots \vee u_{t_2(n)} .$$

From this we conclude as in the proof of (2) that $T_3[t_2(n)] \leqq T_1(n)$. If $t_2(n)$ fails to become infinite with $n$, then $p_*(b, c) = 0$, and (4) is immediate. And when $t_2(n) \to \infty$ with $n$, we may write

$$\frac{t_2(n)}{n} \frac{T_3[t_2(n)]}{t_2(n)} \leqq \frac{T_1(n)}{n},$$

whence on letting $n \to \infty$ the inequality (4) is established.

Turning lastly to the inequality

$$p_*(a, c) \leqq p_*(a, b)p^*(b, c), \tag{5}$$

we may always assume $t_1(n) \leqq T_2(n)$. For if for some $n > 0$ $t_1(n) > T_2(n)$ we should have by Theorem 14

$$b/c \prec u_1 \vee \cdots \vee u_{T_2(n)}/u < u_1 \vee \cdots \vee u_{t_1(n)}/u \prec a/c$$

and $b/c < a/c$ contradicts $a \subset b$, in view of Theorem 3. We may then write

$$a_1 = u_1 \vee \cdots \vee u_{T_2(n)}, \qquad b_1 = u_1 \vee \cdots \vee u_{t_1(n)}, \qquad h_1 = u,$$

$$a_2 = b, \qquad b_2 = a, \qquad h_2 = c,$$

and exclude $a_1b_1h_1 = 0$ as leading to $t_1(n) = 0$ and thus to a triviality, and also observe that $a_2b_2h_2 = a \neq 0$. Now by definition of $t_1(n)$, $T_2(n)$, etc., we have the following relations

$$a_1b_1/h_1 \prec a_2b_2/h_2, \qquad a_1/h_1 \succ a_2/h_2,$$

and thus by Axiom D, a relation obtains which takes the form

$$u_1 \vee \cdots \vee u_{t_1(n)}/u_1 \vee \cdots \vee u_{T_2(n)} \prec a/b;$$

thus, as before, $t_1(n) \leqq t_3[T_2(n)]$. If $T_2(n)$ failed to become infinite with $n$, $p^*(b, c) = 0$, whence by (3) $p^*(a, c) = 0$, so that $p_*(a, c) = 0$, giving (5) trivially. Finally, in the case where $T_2(n) \to \infty$ with $n$, we have

$$\frac{t_1(n)}{n} \leqq \frac{T_2(n)}{n} \frac{t_3[t_2(n)]}{T_2(n)}$$

whence the inequality (5) when $n \to \infty$. Thus the proof of our theorem is complete.

DEFINITION 2. *The eventuality $a/h$ shall be said to be appraisable in the case where $p_*(a, h) = p^*(a, h)$. When $a/h$ is appraisable, the common limit*

$$p(a, h) = p_*(a, h) = p^*(a, h)$$

*shall be called the (numerical) probability of $a/h$ (or of the contingency $a$ on the presumption $h$).*

It may be remarked that if every $\mathfrak{A}$ were *completely* ordered by $\prec$, every eventuality in it would be appraisable.

If $a_1/h$ and $a_2/h$ are appraisable, then the inequality $p(a_1, h_1) < p(a_2, h_2)$

implies the relation $a_1/h_1 < a_2/h_2$ by Theorem 16. But the equation $p(a_1, h_1) = p(a_2, h_2)$ is perfectly consistant with any one of the relations $a_1/h_1 <, >, \approx, \| \; a_2/h_2$. Thus relations of numerical probabilities furnish but a blurred picture of the more fundamental comparisons in probability; a circumstance which is illustrated by the familiar fact that we may have $h(a, h) = 0$ even when $ah \neq 0$.

THEOREM 20. *If $(a_1, \cdots, a_\nu)$ is a $\nu$-scale ($\nu = 1, 2, \cdots$) and if $b$ = sum of $\tau$ distinct elements $a_i$ :*

$$b = a_{i_1} \vee \cdots \vee a_{i_\tau} \; (b = 0 \text{ for } \tau = 0),$$

*then $b/a$ is appraisable, and $p(b, a) = \tau/\nu$. In particular (whenever $a \neq 0$), $0/a$ and $a/a$ are appraisable, and $p(0, a) = 0$, $p(a, a) = 1$.*

In virtue of Theorem 14, to say that $t(n)$ is the maximum $t$ for which $u_1 \vee \cdots \vee u_t/u \prec b/a$ is to say that it is the maximum $t$ such that $t/n \leqq \tau/\nu$, and for such $t(n)$, $\lim t(n)/n = \tau/\nu$. And likewise $\lim T(n)/n = \tau/\nu$, and our theorem is evident.

THEOREM 21. *In order that $a/h$ be appraisable it is necessary and sufficient that, corresponding to any $\epsilon > 0$, there should exist appraisable $a'/h'$ and $a''/h''$ such that*

$$a'/h' \prec a/h \prec a''/h'',$$

$$p(a'', h'') - p(a', h') < \epsilon.$$

The necessity of the condition appears at once on setting $a' = a'' = a$, $h' = h'' = h$.

To prove the sufficiency, it is only necessary to observe that $p(a', h') \leqq p_*(a, h)$: otherwise, by Theorem 16, $a'/h' > a/h$; and similarly, $p^*(a, h) \leqq p(a'', h'')$. Hence $p^*(a, h) - p_*(a, h) < \epsilon$ and this $\epsilon$ being arbitrarily small, $p^*(a, h) = p_*(a, h)$ follows.

An obvious theorem is the following:

THEOREM 22. *If $p^*(a, h) = 0$, $a/h$ will be appraisable. If $p_*(a, h) = 1$, $a/h$ will be appraisable.*

THEOREM 23. *If $a/h$ is appraisable, then $\sim a/h$ will be appraisable, and $p(a, h) + p(\sim a, h) = 1$ will hold.*

For on applying Theorem 17, first directly, then with $a$ and $\sim a$ interchanged, we obtain

$$p_*(a, h) + p^*(\sim a, h) = 1,$$

$$p^*(a, h) + p_*(\sim a, h) = 1$$

which in conjunction with the hypothesized relation $p_*(a, h) = p^*(a, h)$, make the theorem evident.

THEOREM 24. *If $a_1 a_2 h = 0$ and $a_1/h$ and $a_2/h$ are both appraisable, then $a_1 \vee a_2/h$ will be appraisable, and*

$$p(a_1 \vee a_2, h) = p(a_1, h) + p(a_2, h).$$

This is the *principle of total probability*; it may be extended to any *finite* number of disjoint summands. Its truth is an immediate consequence of Theorem 18.

THEOREM 25. *If $a \subset b \subset c$ and if $a/b$ and $b/c$ are appraisable, then $a/c$ will be appraisable and $p(a, c) = p(a, b)p(b, c)$ will hold.*

This follows immediately from Theorem 19.

THEOREM 26. *If $a \subset b \subset c$ and $a/c$ and $b/c$ are appraisable, and lastly if $p(b, c) \neq 0$, then $a/b$ will be appraisable, and $p(a, c) = p(a, b)p(b, c)$ will hold.*

Setting $p_*(b, c) = p^*(b, c) = p(b, c)$ and $p_*(a, c) = p^*(a, c) = p(a, c)$ in Theorem 19, we have

$$\begin{aligned} p_*(a, b)p(b, c) &\leqq p(a, c) \leqq p_*(a, b)p(b, c), \\ p^*(a, b)p(b, c) &\leqq p(a, c) \leqq p^*(a, b)p(b, c) \end{aligned} \tag{6}$$

whence $p^*(a, b)p(b, c) \leqq p_*(a, b)p(b, c)$, and the proof is completed by dividing by $p(b, c)$.

THEOREM 27. *If $a \subset b \subset c$ and if $a/b$ and $b/c$ are appraisable, then $a \sim b/c$ is appraisable, and $p(a \sim b, c) = p(a, c) - p(b, c)$.*

First suppose that $p(b, c) = 0$. The above proof of (6) may be applied, whence it follows that $p(a, c) = 0$. On the other hand $b \sim a \subset b$; hence $b \sim a/c \prec b/c$ (Theorem 3), so that by Theorem 16 $0 \leqq p_*(b \sim a, c) \leqq p^*(b \sim a, c) \leqq p(b, c) = 0$. Thus $b \sim a/c$ is appraisable, and $p(b \sim a, c) = 0 = p(a, c) - p(b, c)$.

Next, let $p(b, c) \neq 0$. Then by Theorem 26, $a/b$ is appraisable, and hence by Theorem 23 $\sim a/b$ is appraisable and $p(\sim a, b) = 1 - p(a, b)$. Theorem 25 with $b \sim a \subset b \subset c$ replacing $a \subset b \subset c$ gives the appraisability of $b \sim a/c$ and we have

$$\begin{aligned} p(b \sim a, c) = p(b \sim a, b)p(b, c) = p(\sim a, b)p(b, c) &= [1 - p(a, b)]p(b, c) \\ &= p(b, c) - p(a, b)p(b, c) = p(b, c) - p(a, c). \end{aligned}$$

THEOREM 28. *Let $a_1$, $a_2$, $h$ be any propositions such that $a_1 \subset h$, $a_2 \subset h$, and that $a_1/h$ and $a_2/h$ are appraisable. Then $a_1 \vee a_2/h$ will be appraisable if and only if $a_1a_2/h$ is appraisable, and*

$$p(a_1, h) + p(a_2, h) = p(a_1 \vee a_2, h) + p(a_1a_2, h).$$

Firstly, let $a_1a_2/h$ be appraisable. If $p(a_1, h) = 0$ and $p(a_2, h) = 0$, it would follow by Theorem 18 that $p^*(a_1 \vee a_2, h) = 0$, whence the appraisability of $a_1 \vee a_2/h$ (Theorem 22). In the contrary case, we should have, let us say, $p(a_1, h) > 0$. Then by Theorem 27 applied with $a = a_1a_2$, $b = a_1$, it would follow (since $a_1 \sim (a_1a_2) = a_1 \sim a_2$) that $a_1 \sim a_2/h$ is appraisable, and that $p(a_1 \sim a_2, h) = p(a_1, h) - p(a_1a_2, h)$. Now we have $a_1 \vee a_2/h = a_1 \sim a_2 \vee a_2/h$, $(a_1 \sim a_2)a_2h = 0$; and hence by Theorem 24, $a_1 \vee a_2/h$ is appraisable, and $p(a_1 \vee a_2, h) = p(a_1 \sim a_2, h) + p(a_2, h) = p(a_1, h) - p(a_1a_2, h) + p(a_2, h)$; and the first half of our theorem is established.

Secondly, let $a_1 \vee a_2/h$ be appraisable. Theorem 23 then establishes the appraisability of $\sim(a_1 \vee a_2)/h = \sim a_1 \sim a_2/h$, $\sim a_1/h$, and $\sim a_2/h$, and gives the relations

$$p(\sim a_1 \sim a_2 , h) = 1 - p(a_1 \vee a_2 , h),$$

$$p(\sim a_1 , h) = 1 - p(a_1 , h), \qquad p(\sim a_2 , h) = 1 - p(a_2 , h).$$

Now the established half of the present theorem yields the appraisability of $\sim a_1 \sim a_2/h$, and the equation

$$p(\sim a_1 , h) + p(\sim a_1 , h) = p(\sim a_1 \vee \sim a_2 , h) + p(\sim a_1 \sim a_2 , h).$$

A second application of Theorem 23 gives us the appraisability of $\sim(\sim a_1 \vee \sim a_2)/h = a_1 a_2/h$, and also that $p(a_1 a_2 , h) = 1 - p(\sim a_1 \vee \sim a_2 , h)$. On combining these equations, the desired conclusion is established.

We thus reach the conclusion of this section, and with it our final goal, having defined the numerical probability and established its conventional properties: In virtue of the Theorem of Representation of Boolean Algebras of Garrett Birkhoff and M. H. Stone[14] our underlying Boolean ring $\mathfrak{A}$ corresponds with a closed system of subclasses of a certain $E$, in a subsystem of which a (restrictedly) additive set function $p_H(A) = p(a, h)$ is defined which satisfies the Axioms of Kolmogoroff,[15] from which the classical theory of probability may be derived.

Our proposed task is thus accomplished.

The relation with frequency in sequences will be studied in a forthcoming paper.

COLUMBIA UNIVERSITY.

[14] For full references and discussion, see M. H. Stone, The Representation of Boolean Algebras, Bull. Amer. Math. Soc., Vol. 44, Dec. 1938, pp. 807–816.

[15] A. Kolmogoroff, Grundbegriffe der Wahrscheinlichkeits Rechnung. Ergebnisse der Mathematik, B2. (Berlin 1933.)

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# GENERAL ERGODIC THEOREMS

BY L. ALAOGLU AND G. BIRKHOFF

(Received February 3, 1940)

## 1. Introduction

Let $B$ be any group or semi-group of linear operators $T_\sigma$ on a Banach space $E$. By a (mean) "ergodic theorem," we mean a theorem asserting the convergence of the means $\sum \lambda_\sigma x T_\sigma$ of the transforms of an element $x$ of $E$ to a fix-point.

We prove below two main ergodic theorems. First, *if* $| xT_\sigma | \leqq | x |$ *identically, if* $E$ *is uniformly convex and has a unit sphere without "sharp edges" (as with the spaces* $L_p$ *and* $l_p$ $(p > 1)$*), then every* $x$ *is ergodic.* Second, *if* $G$ *is "ergodic" in a sense including n-parameter Abelian groups, Lie groups with nilpotent Lie algebras, and others, and if there is a convex compact set containing the transforms of* $x$*, then* $x$ *is ergodic.*

The case $G$ cyclic, $| xT_\sigma | = | x |$ identically, and $E$ Hilbert space was handled by von Neumann [9][1] in 1931, while G. D. Birkhoff [3] obtained stronger results in a special case important for dynamics. Various authors (cf. [4], [10], [8], and [12]) tried to prove somewhat later similar results for more general $E$ and $T_\sigma$, but still with cyclic $G$. Notably, the identity $| xT_\sigma | = | x |$ was weakened to $| xT_\sigma | \leqq K \cdot | x |$, and $E$ was allowed to be any space $L_p$.

More recently, N. Dunford [7] and Wiener [11] treated the case where $G$ was an $n$-parameter Abelian group. Dunford showed that the methods of Riesz, Kakutani and Yosida applied to any such $G$; and our second main theorem shows that one need in fact merely assume that $G$ has a sequence of "nearly invariant" measures, and makes certain simplifications.

However, this still imposes a strong restriction on $G$, which is completely absent in our first main theorem. Our proof here depends on generalizing Wiener's argument (which was for $G$ an $n$-parameter Abelian group, $E$ Hilbert space, and $| xT_\sigma | \leqq | x |$) to the case $E$ uniformly convex [5]. Further, one needs a new concept of limit and a new argument assuming that the unit sphere in $E$ has no sharp edges, to remove the restriction on $G$; these are in [6].

The results of the present paper were sketched in [1]; we propose to continue these researches in another paper.

## I. GENERALITIES

### 2. Generalized convergence

The elementary notion of the convergence of a sequence of (real) numbers $x_1, x_2, x_3, \cdots$ to a limit, can be generalized in two ways. One can replace

[1] Numbers in square brackets refer to the bibliography.

numbers by more general elements, and one can replace sequences by more general orderings.

We shall replace numbers by points of any topological space satisfying Hausdorff's axioms.[2] And we shall allow these points to be ordered by any *transitive* relation. More precisely, we shall mean by a "generalized sequence," any set $\{x_\alpha\}$ of terms $x_\alpha$ between which a transitive relation $x_\alpha < x_\beta$ (read, $x_\beta$ is a *successor* of $x_\alpha$) is defined. We allow the same element to occur repeatedly in the sequence.

Relative to these concepts, one can formulate ([6])

DEFINITION 1: *A generalized sequence $\{x_\alpha\}$ of points of a topological space "converges" to a limit $a$ (in symbols, $x_\alpha \to a$), if and only if, given any neighborhood $U(a)$ of $a$, each term of the sequence has a successor, all of whose successors lie in $U(a)$.*

## 3. Examples

This definition obviously contains as a special case, the usual definition of convergence for an ordinary sequence of points of a topological space. More generally,

THEOREM 1: *Definition* 1 *specializes to the usual definition of Moore-Smith convergence in the case of generalized sequences having the Moore-Smith property.*[3]

For this definition is, that to each $U(a)$ corresponds a term $x_\alpha$, all of whose successors lie in $U(a)$. This is clearly implied by Definition 1. And conversely, if it is satisfied, then given $U(a)$, there exists an $x_\alpha$ all of whose successors lie in $U(a)$—hence to each $x_\beta$ a successor $x_\gamma$ (namely, any common successor of $x_\alpha$ and $x_\beta$), all of whose successors lie in $U(a)$. This implies Definition 1, completing the proof.

It is a corollary that Definition 1 specializes to the usual concept of convergence for transfinite sequences.

## 4. Properties

In discussing the properties of generalized convergence, we shall be guided by the known results of Fréchet and others on the properties of ordinary sequences.

First, we define a *subsequence* of a generalized sequence $\{x_\alpha\}$, as a subset containing at least one successor of every term in the original sequence. This is usually called a "confinal subset"; since ordering is transitive, evidently any subsequence of a subsequence of a given generalized sequence, is itself a subsequence of the original sequence. It is also evident that

(1) If a generalized sequence $\{x_\alpha\}$ converges to a given limit $a$, then so do all its subsequences.

---

[2] That is, a space of points having *neighborhoods* satisfying axioms (A)–(C) and (6) of F. Hausdorff's *Mengenlehre*, 3d ed. (1935), pp. 228–9. Such spaces are usually called "Hausdorff spaces."

[3] Namely, the property that any two terms have a common successor. Cf. E. H. Moore and H. L. Smith, *A general theory of limits*, Am. Jour. 44 (1922), pp. 102–21, and G. Birkhoff, *Moore-Smith convergence in general topology*, these Annals 38 (1937), pp. 39–57.

(2) If all the terms of a generalized sequence $\{x_\alpha\}$ are the same point $a$, then $x_\alpha \to a$.

(3) If $x_\alpha \to a$ and $x_\alpha \to b$, then $a = b$. (In words, a generalized sequence converges to at most one limit.) The proof of this depends on Hausdorff's separation axiom (6).

(4) Unless $x_\alpha \to a$, there exists a neighborhood $U(a)$ and element $x_\alpha$, each of whose successors $x_\beta$ has a successor not in $U(a)$. Hence if we delete from $\{x_\alpha\}$ the successors of this $x_\alpha$ in $U(a)$, we will get a subsequence $\{y_\alpha\}$, no subsequence of which converges to $a$. Turning this around, we conclude that if every subsequence of a generalized sequence $\{x_\alpha\}$ contains a sub-subsequence converging to $a$, then $x_\alpha \to a$.

Properties (1)–(4) are generalizations of the defining properties of $L^*$-spaces in the sense of Fréchet.[4] We note also that if a generalized sequence $\{x_\alpha\}$ is the set-theoretic sum of a finite number of subsequences each converging to the limit $a$, then $x_\alpha \to a$. However, as we shall not need them, we shall not discuss in full the abstract properties of generalized convergence—in spite of the interest of this topic for general topology.

## 5. Further properties

We assume the reader to be acquainted with the definition and elementary properties of Banach spaces,[5] and wish simply to point out

(5) In a Banach space, $x_\alpha \to a$ and $y_\alpha \to b$ imply $x_\alpha + y_\alpha \to a + b$; while $x_\alpha \to a$ and $\lambda_\alpha \to \lambda$ imply $\lambda_\alpha x_\alpha \to \lambda a$.

In fact, given $\epsilon > 0$, every $\alpha$ has a successor $\beta$ such that all successors of $x_\beta$ satisfy $|x_\nu - a| < \frac{1}{2}\epsilon$, while $\beta$ has a successor $\gamma$ such that all successors of $y_\gamma$ satisfy $|y_\nu - b| < \frac{1}{2}\epsilon$. All successors $\nu$ of $\gamma$ will satisfy $|(x_\nu + y_\nu) - (a + b)| < \epsilon$. This proves $x_\alpha + y_\alpha \to a + b$; the proof that $\lambda_\alpha x_\alpha \to \lambda a$ is identical.

Actually, condition (5) holds in any topological space, and analogous conditions hold more generally in any topological algebra.[6] Indeed, define a function $f(x_1, \cdots, x_n)$ of variables $x_i$ in topological spaces $E_i$, with values in a topological space $F$, to be *continuous* whenever, given a neighborhood $V$ of $f(a_1, \cdots, a_n)$ neighborhoods $U_i$ of the $a_i$ can be found so small that $f(U_1, \cdots, U_n) \leqq V$ in the sense of the calculus of complexes.[7] Then for continuous functions, $x_\alpha' \to a_1$, $\cdots$, $x_\alpha^n \to a_n$, imply that

$$f(x_\alpha', \cdots, x_\alpha^n) \to f(a_1, \cdots, a_n)$$

in the sense of generalized convergence.

---

[4] Cf. P. Urysohn, *Sur les classes* (L) *de M. Fréchet*, Enseignement Math. 25 (1926), pp. 77–83.

[5] Or "espaces du type $(B)$" in the terminology of Banach [2]. These are simply complete metric linear spaces in the sense of Hausdorff, op. cit.

[6] Topological group, ring, field, etc. Cf. D. van Dantzig, *Zur topologischen Algebra*, Math. Annalen 107 (1933), pp. 587–626.

[7] That is, $f(U_1, \cdots, U_n)$ signifies the set of $f(x_1, \cdots, x_n)$ with $x_1 \in U_1, \cdots, x_n \in U_n$.

## 6. Definition of ergodicity

Throughout the sequel, we shall be concerned with groups or semi-groups[8] $G$ of linear operators $T_g$ of bounded modulus[8] on a Banach space $E$.

By a "fix-point" of $E$, we mean an element $a$ such that $aT_g = a$ for all $T_g \in G$. It is clear by (5) that the fix-points form a closed linear subspace $F$ of $E$; in fact this holds for continuous linear operators on any topological linear space, by the argument of §5.

Elements not fix-points will have in general many transforms; the convex hull of the set of transforms $xT_g$ of any element $x$, consisting of the weighted *means* or *averages*

$$\sum \lambda_g x T_g \qquad (\lambda_g \geqq 0, \sum \lambda_g = 1)$$

of its transforms, will play an important role below.

By the *successors* of a mean of transforms of $x$, we signify the means of *its* transforms. Since

$$\sum \mu_h (\sum \lambda_g x T_g) T_h = \sum_{g,h} \mu_h \lambda_g x (T_g T_h) \qquad (\mu_h \lambda_g \geqq 0, \sum_{g,h} \mu_h \lambda_g = 1)$$

these are means of transforms of $x$; for the same reason, the predecessor-successor relation is transitive. Hence it makes sense to state, after these preliminaries,

DEFINITION 2: *An element $x$ of $E$ is "ergodic" if and only if the means of its transforms "converge" in the sense of Definition 1 to a fix-point. The set $\{T_g\}$ is ergodic if and only if all elements of $E$ are ergodic.*

## 7. Equivalent conditions

We shall now establish the equivalence with Definition 2 of three other definitions of ergodicity.

First, let $G^*$ be the convex hull of $G$: the set of means $T_h^* = \sum \lambda_g T_g$ of the operators of $G$. Then $G^*$ will be itself a semi-group with the same fix-points and the same upper bound $K$ to the moduli of its operators, as $G$. Now order its elements by the convention that the "successors" of any $T_h^*$ are its right-multiples $T_h^* T_i^*$. Definition 2 is obviously simply the condition that the $xT_h^*$ converge under this ordering. Again,

LEMMA 1: *For $x$ to be ergodic, it is necessary and sufficient that, given $T_g$, there exist a mean of transforms of $xT_g$ arbitrarily near $a$,* where $a$ is a FIXED POINT.

---

[8] By a "semi-group," we mean simply a set of operators which contains with any two operators, their product (following the usage of H. S. Vandiver and others). By the "modulus" (or "norm") of an operator $T$, we mean

$$\| T \| = \sup_{x \neq y} | xT - yT | / | x - y | = \sup_{|x|=1} | xT |$$

these being the same for linear operators on a Banach space. Thus our assumption is that for some finite $K$ independent of $T_g$, the inequality $| xT_g | \leqq K \cdot | x |$ holds identically.

PROOF: The condition is obviously necessary. To show its sufficiency, note that given $\epsilon > 0$ and $T^* = \sum_{k=1}^{n} \lambda_k x T_k$, we can by the condition choose in succession $T_1^*, \cdots, T_n^*$ such that $| xT_kT_1^* \cdots T_k^* - a | < \epsilon/K$. From this we can infer

$$| xT^* T_1^* \cdots T_n^* - a | \leqq \sum_{k=1}^{n} | (xT_kT_1^* \cdots T_k^* - a)T_{k+1}^* \cdots T_n^* | \cdot (\epsilon/K)K$$

which is the first alternative condition for ergodicity above.

THEOREM 2: *An element $x$ is ergodic if and only if the closure $K(x)$ of the convex hull of the set of its transforms contains one and only one fix-point.*

PROOF: If $x$ is ergodic, then the means of its transforms converge to a limit fix-point $a$, which is clearly in $K(x)$; moreover if $b$ were any second fix-point in $K(x)$, then one could find $T^* \in G^*$ such that $| xT^* - b | < | a - b |/2K$; whence for all $T_i^*$, $| xT^*T_i^* - b | < | a - b |/2$—and so $| xT^*T_i^* - a | > | a - b |/2$, contradicting ergodicity.

Conversely, if the condition is satisfied, then for any $T_g \in G$, $K(xT_g) \leqq K(x)$ will contain the unique fix-point $a$ of $K(x)$ for its fix-point. Consequently there will be $xT_gT^*$ arbitrarily near $a$, implying ergodicity by Lemma 1.

## 8. Digression on ergodic subspaces

It is easy to prove that under the hypotheses of §6, which we shall always assume,

THEOREM 3: *The ergodic elements of $E$ form a closed, invariant linear subspace.*

PROOF: By condition (5) of §5, they form a subspace. Moreover this subspace is invariant since every $xT_g$ is ergodic if $x$ is, by direct application of Definition 2. In fact, these remarks hold for general linear operators on a topological linear space.

The proof of closure involves the assumption of bounded modulus: if $x_i \to x$ and the means of the transforms of the $x_i$ tend to limits $a_i$, then the $a_i$ converge to a limit $a$, and the means of the transforms of $x$ converge to this $a$; in fact

$$| xT^* - a | \leqq | xT^* - x_iT^* | + | x_iT^* - a_i | + | a_i - a |,$$

and this can be made arbitrarily small by choosing $i$ large enough, and then $T^*$ suitably.

We note that this subspace—which we shall call the "ergodic subspace"—is the sum of the subspace $F$ of fix-points, and the subspace $N$ of elements, the means of whose transforms tend to zero. Indeed, $F \cap N = 0$: the ergodic subspace is the direct sum of $F$ and $N$. One might call $N$ the "dissipative subspace."

REMARK: If the means of the transforms of an element $x$ converge to a limit $a$, then $a$ is a fix-point and so $x$ is ergodic. For, given $\epsilon > 0$, we can so choose $T^* \in G^*$ that $| xT^* - a | < \epsilon$ and $| xT^*T_g - a | < \epsilon$ for all $T_g$ of $G$. This implies $| xT^*T_g - aT_g | < K\epsilon$ by the modulus restriction, and so $| aT_g - a | < (K + 1)\epsilon$ for all $T_g$ and $\epsilon > 0$, whence $| aT_g - a | = 0$ and $a$ is a fix-point.

But now if $X$ is any closed subspace of a Banach space $E$, one can define a quotient-space $E/X$ as follows. The elements of $E/X$ are the different hyperplanes $a + X$ parallel to $X$; the norm of $a + X$ is $\inf_{x \in X} |a + x|$. We omit the proof that this is a Banach space; the technique is straightforward and elementary.[9]

Moreover any linear operator leaving $X$ *invariant* is half-reduced by it (in the language of representation theory), and is a linear operator of no greater modulus on $E/X$. This suggests that if the ergodic subspace $X$ is a proper subspace $(0 < X < E)$, it may be interesting to consider the operation of $G$ on the quotient-space $E/X$.

For example, let $G$ be the cyclic group of linear isometries $T^\lambda$: $f(x) \to f(x + \lambda)$ on the space $(L)$ of Lebesgue-integrable functions on the interval $(-\infty, +\infty)$. Obviously any fix-point is a constant function $f(x) = c$, and so 0 is the only fix-point. The ergodic subspace is thus the $N$ defined above: it includes all functions with $\int_{-\infty}^{+\infty} f(x)\,dx = 0$, as can be proved quite easily. The means of the transforms of other functions do not converge even weakly;[10] hence the ergodic subspace $X$ includes no other functions. In this case the quotient-space $E/X$ is a line, all of whose points are fix-points under the group.

## II. Fundamental Theorems

### 9. Survey

Let us now go back to the basic situation of §6: a semi-group $G$ of linear operators $T_g$ of bounded moduli, operating on a Banach space $E$. We wish to find sufficient conditions for the ergodicity of all elements of $E$, realizing from the example of §8 that some additional restrictions must be imposed.

We shall prove two groups of theorems, based on entirely different arguments. In theorems of the first group, purely metric arguments will be used. No restriction will be put on $G$ as an abstract group, but the moduli of the $T_g$ will be required to be at most one, and $E$ will be assumed to be uniformly convex. These theorems are the most original, and the furthest from the classical ergodic theorems of G. D. Birkhoff [3] and J. von Neumann [9].

The theorems of the second group depend on a more sophisticated argument, due primarily to F. Riesz [10]. Results applying to all Banach spaces will be proved, and no restriction beyond that of bounded moduli will be put on the $T_g$. On the other hand, $G$ will be greatly restricted—although not enough to exclude any cases treated by other authors (notably, cyclic and $n$-parameter Abelian groups).

---

[9] This construction (apart from metric considerations) can be extended to closed subspaces of topological linear spaces and closed normal subspaces of topological groups; cf. L. Pontrjagin, *Topological groups*, Princeton, 1939, p. 59, Thm. 11. The hypothesis that $X$ is closed is essential; the only *non-trivial point is* the proof of completeness.

[10] Cf. G. Birkhoff, *Lattice theory*, New York, 1940, Chap. IX, for the construction, and note Theorem 7 below for the equivalence of our notion of ergodicity with that used there.

Finally, we shall show by counterexamples that one cannot drop either set of restrictions: if one allows the moduli of the $T_g$ to exceed one, one *must* restrict $G$; and if one allows $G$ to be arbitrary, one *must* restrict both $E$ and the moduli of the $T_g$.

## 10. The minimal method

We shall now develop the first group of theorems. Their proof depends ultimately on the purely geometrical

LEMMA: *Let $K$ be any closed convex subset of any uniformly convex Banach space. Then there is one and only one point of $K$ of least norm.*[11]

PROOF: Let $M$ be the infimum of the norms of the elements of $K$. Evidently $|a| = |b| = M$ would imply $|\frac{1}{2}(a+b)| < M$ if $a \neq b$, which is impossible; hence there is only one such point. To prove there *is* such a point, choose $\{x_n\}$ from $K$ so that $|x_{n+1}| \leqq |x_n| < M + 1/n$. Then $|x_{n+k} - x_n| > \epsilon$ implies by definition $|\frac{1}{2}(x_n + x_{n+k})| \leqq M + 1/n - \delta$—and hence $n \leqq 1/\delta$ since $\frac{1}{2}(x_n + x_{n+k})$ is in $K$. Therefore $\{x_n\}$ is *convergent*; its limit $a$ will evidently be a point in $K$ of norm $M$, completing the proof.

It is a corollary that if $K$ is carried into itself by every $T_g$—which is the case if $K$ is the closed convex hull $K(x)$ of the set of transforms of some $x$—then this point is carried by every $T_g$ into a point of $K$ of no greater norm, and so into itself. Consequently

THEOREM 4:[12] *If the $T_g$ have moduli bounded by one, and $E$ is uniformly convex, then the closed convex hull $K(x)$ of the set of transforms of any element $x$ contains at least one fix-point—namely, the element of $K(x)$ nearest* 0.

## 11. Abelian case

Comparing Theorem 4 with Theorem 2, we see that if $K(x)$ contains *at most* one fix-point, then we have ergodicity. We shall now show that this is true in at least two important cases.

First we consider the case that $G$ is Abelian; we do not at present assume that $E$ is uniformly convex, or that the moduli of the $T_g$ are at most one.

LEMMA 1: *If $G$ is an Abelian semi-group, then the means of the transforms of any $x$ have the property of Moore-Smith.*

PROOF: Given $\sum \lambda_g x T_g$ and $\sum \mu_h x T_h$, these means have the common successor $\sum \lambda_g \mu_h x T_h T_g = \sum \mu_h \lambda_g x T_g T_h$.

---

[11] By a "uniformly convex" Banach space, we mean (J. A. Clarkson, *Uniformly convex spaces*, Trans. Am. Math. Soc. 40 (1936), pp. 396–414), one in which, given $\epsilon > 0$, there exists $\delta > 0$ so small that $|y| \leqq |x| \leqq 1$ and $|x - y| \geqq \epsilon$ implies $|\frac{1}{2}(x+y)| \leqq |x| - \delta$. That Hilbert space is uniformly convex follows from

$$|x+y|^2 + |x-y|^2 = 2|x|^2 + 2|y|^2 \leqq 4|x|^2$$

whence $|\frac{1}{2}(x+y)| \leqq (|x| - \frac{1}{4}|x-y|^2)^{\frac{1}{2}}$. The proof below is in [5] and [6].

[12] This point is really the crux of Wiener's argument [11]; by lifting it from its context, we get much more from it.

REMARK: The conclusion of Lemma 1 also holds in certain other cases—for instance, if $G$ (or $G^*$) has a right-zero (or operator $T_0$ satisfying $T_\alpha T_0 = T_0$ for all $T_\alpha$); it also holds if $G$ is a finite group—and approximately if $G$ has an invariant finite measure (cf. §20). It does *not* hold for all groups—cf. Example 2 of §13.

LEMMA 2: *If the means of the transforms of $x$ have the property of Moore-Smith, then $K(x)$ contains at most one fix-point.*

PROOF: Suppose two distinct fix-points $a$ and $b$; choose means $y$ and $z$ of transforms of $x$ whose distances from $a$ and $b$ are at most $|a - b|/4K$. There would exist a mean $u$ of transforms of $y$ which was also a mean of transforms of $z$. But since $a$ and $b$ were fix-points and our operators have moduli at most $K$, this would imply $|u - a| \leqq \frac{1}{4}|a - b|$ and $|u - b| \leqq \frac{1}{4}|a - b|$, which is absurd.

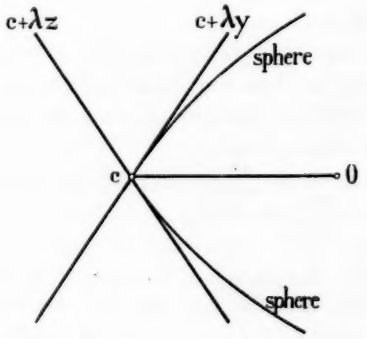
Combining Lemmas 1–2 with Theorem 4, we get a direct extension of Wiener's mean ergodic theorem,

THEOREM 5: *If $G$ is Abelian, $E$ is uniformly convex, and the moduli of the $T_\alpha$ at most one, then every element of $E$ is ergodic.*

## 12. General case

In the general case, we have to use an entirely new argument, involving an additional restriction on $E$. In fact, as we shall see in §13, Theorem 5 does not remain true if $G$ is unrestricted.

Suppose indeed that under the hypotheses of Theorem 4 (we could discard uniform convexity), the closed convex hull $K(x)$ of the set of transforms of some $x$ contains *two* fix-points $a$ and $b$. Then $K(x - a) = K(x) - a$ and $K(x - b) = K(x) - b$ both contain 0: Set $x - a = y$, $x - b = z$, and $a - b = z - y = c$; $c$ will be a fix-point.




By construction, we can find for any real $\lambda$ and $\epsilon > 0$, an element $T^*$ in the convex hull $G^*$ of $G$ satisfying $|\lambda y T^*| < \epsilon$; hence for any fixed $\lambda$, $(c + \lambda y)T^*$ can be found arbitrarily near $c$. But for all $T^*$, $|(c + \lambda y)T^*| \leqq |c + \lambda y|$ by hypothesis; hence we conclude $|c + \lambda y| \leqq |c|$ for all $\lambda$, and similarly $|c + \lambda z| = |-c + (-\lambda)z| \leqq |c|$ for all $\lambda$.

In the $(y, z)$-plane—which includes $c = z - y$,—we get thus the following picture: the "sphere" of radius $|c|$ about the origin lies inside both the lines $c + \lambda y$ and $c + \lambda z$. That is, it has a *sharp edge*. By similarity, we get as a corollary the precise

LEMMA: *If some $K(x)$ contains more than one fix-point, then the "circle" where the unit sphere intersects some plane has an angle in it: it is not a smooth curve.*[13]

Combining this lemma with Theorems 2 and 4, we get

THEOREM 6: *If the moduli of the $T_g$ are at most one, if $E$ is uniformly convex and its unit sphere has no sharp edges, then every element of $E$ is ergodic.*[14]

COROLLARY: *If $E$ is a space $L_p$ or $l_p$ $(p > 1)$, or Euclidean space of any dimension, and the moduli of the $T_g$ are at most one, then every element of $E$ is ergodic.*

## 13. Phenomenon of multiple fix-points

We shall now exhibit instances where the closed convex hull $K(x)$ of the set of transforms of some $x$ contains many fix-points. By §11, this phenomenon occurs only if $G$ is non-commutative, and so has not occurred in the cases discussed in the literature on ergodic theory.[15] Also, by §12, if the moduli of the $T_g$ are at most one, it occurs only if the unit sphere of $E$ has a sharp edge.

EXAMPLE 1: Let $E$ be the $(\xi, \eta)$-plane, and let $G$ consist of the two transformations

$$T: \quad (\xi, \eta) \to (\xi, 0) \qquad T': \quad (\xi, \eta) \to (\xi + \tfrac{1}{2^0}\eta, 0).$$

Then if $x = (0, 1)$, $K(x)$ contains both $xT = (0, 0)$ and $xT' = (\frac{1}{2^0}, 0)$, which are distinct fix-points.

Relative to the norm $|(\xi, \eta)| = |\xi| + |\eta| + \frac{1}{10^0}(\xi^2 + \eta^2)^{\frac{1}{2}}$, $E$ is uniformly convex, and $||T|| = ||T'|| = 1$. This shows that one cannot drop either the requirement that $G$ be commutative in Theorem 5, nor the requirement that the unit sphere have no sharp edge in Theorem 6. Moreover the moduli of $T$ and $T'$ are bounded if we use the Euclidean norm $(\xi^2 + \eta^2)^{\frac{1}{2}}$; hence the restriction to the case of moduli at most one in Theorem 6 is also necessary.

EXAMPLE 2: Let $E$ be the space $(B)$ of bounded real functions $f(x)$ with the norm $||f|| = \sup f(x)$, and let $G$ be the group of isometries of $E$ induced by permutations of the variable $x$. Define $f_0(x)$ to have the value one if $x > 0$, and zero if $x \leqq 0$. Then $K(f_0)$ includes all constant functions with values

---

[13] This is equivalent to the condition that the unit sphere have two tangent planes at some point. J. von Neumann remarked to the authors that this is equivalent to the condition that the unit sphere in the conjugate space contain a straight line segment—hence the second condition on $E$ in Theorem 6 is practically dual to the condition of uniform convexity.

[14] This is a direct generalization of the main theorem in §4 of G. Birkhoff [6].

[15] However, J. von Neumann observed in his researches on almost periodic functions on groups (*Almost periodic functions in a group. I*, Trans. Am. Math. Soc. 36 (1934), pp. 445–92; esp. Thm. 6, p. 451) a logical connection between the uniqueness of the mean of a.p. functions, and commutativity.

between 0 and 1, and these are fix-points under $G$; we leave the details to the reader.[16]

This shows that Lemmas 1–2 of §11 do not either of them hold even for all *groups* (as contrasted with semi-groups). However, if $E$ is finite-dimensional and $G$ is a group, then the closure $\bar{G}$ of $G$ is compact and defines the same $K(x)$ as $G$. Moreover $\bar{G}$ possesses an invariant measure,[17] and so by Theorem 9 infra, every element of $E$ is ergodic.

This leaves undecided the question of whether $G$ is a *group*, the $T_g$ of moduli at most one, and $E$ uniformly convex, then every element of $E$ is ergodic.

## 14. Ergodic groups

We now come to the method used by F. Riesz, and independently by S. Kakutani and K. Yosida.[18] This method is less elegant than the arguments used in §§10–12, in that it assumes some rather technical properties of Banach spaces (notably, the equivalence of weak closure with metric closure for convex sets, and the weak compactness of various subsets of certain spaces), and a knowledge of group measures. Thus it imposes very stringent non-algebraic restrictions on $G$; on the other hand, it applies to the space $(L)$ and other spaces which are not uniformly convex, and weakens[19] the condition $\| T_g \| \leqq 1$ to $\| T_g \| \leqq K$.

The discussion will bring out the equivalence between Definition 2 of "convergence" for the means of the transforms of an element, and the (measure-theoretic) concept for this used in the literature. It naturally begins with

DEFINITION 3: *An abstract group or semi-group $G$ is "ergodic" when it possesses a sequence of measures $\mu_n(V)$ satisfying* (i) $\mu_n(G) = 1$ *for all $n$, and* (ii) *given $T_a$ in $G$ and $\epsilon > 0$, $N$ exists so large that $n > N$ implies*[20]

$$| \mu_n(VT_a) - \mu_n(V) | + | \mu_n(T_aV) - \mu_n(V) | < \epsilon.$$

Such a sequence of measures may be characterized roughly by the property of *approaching* invariance under group translations—or of being "nearly invariant."

## 15. Examples

The $n^{\text{th}}$ means in the cyclic case, and the more general means used by Dunford [7] and Wiener [11] for $r$-parameter Abelian groups and semi-groups, furnish

[16] This is a simpler form of the measure-paradox of Hausdorff. Incidentally, the existence of such a paradox is related to the non-commutativity of the group of rigid motions of the domain (cf. J. von Neumann, *Zur allgemeinen Theorie des Masses*, Fund. Math. 13 (1929), pp. 73–116).

[17] A. Haar, *Der Massbegriff in der Theorie der kont. Gruppen*, Annals of Math. 34 (1933), pp. 147–69; also L. Pontrjagin, op. cit., p. 91.

[18] Cf. [10], [12], [8]. The generalization given to $E$ over F. Riesz is perhaps simpler than that of Kakutani and Yosida; cf. G. Birkhoff's *Lattice theory*, Chap. IX.

[19] The latter generalization is however illusory: we can set up a new "unit sphere," the convex hull of the old one and its transforms under $G$. This will give an equivalent topology, and make the moduli of the $T_g$ at most one.

[20] It is understood tacitly that the class of measurable sets $V$ is independent of $n$, and that if $V$ is measurable, then so are $VT_a$ and $T_aV$. (These are the sets of $xT_a$ $(x \in V)$ and $T_ax$ $(x \in V)$.

elementary examples of such sequences of measures. In these cases, $\mu_n(V)$ is the proportion of the cube $S_n$ of elements $T_1^{\lambda_1} \cdots T_r^{\lambda_r}$ satisfying $0 \leqq \lambda_1 \leqq n, \cdots, 0 \leqq \lambda_r \leqq n$, occupied by $V$.

Another example is furnished by group measures. For instance, if $G$ is a compact Lie group (or compact separable topological group!), then[17] $G$ has a measure $\mu(V)$ satisfying $\mu(VT_a) = \mu(T_aV) = \mu(V)$ for all $T_a$, and $\mu(G) = 1$. Here we need only set $\mu_n(V) = \mu(V)$ for all $n$; in fact, the arguments given below are trivial in this case.

One can also show that any Lie group whose Lie algebra is *nilpotent* is ergodic; we will illustrate the method by the case of the three-parameter group whose generators $X$, $Y$, $Z$ satisfy $[XY] = Z$ and $[XZ] = [YZ] = 0$. From such identities as

$$X^\lambda Y^\mu Z^\nu X^\alpha = X^{\lambda+\alpha} Y^\mu Z^{\nu+\alpha\mu}$$

one infers that (1) volume in the $(\lambda, \mu, \nu)$-space is invariant under group-translations, while (2) the set $S_n$ of points satisfying $0 < \lambda, \mu < n$, $0 < \nu < n^3$ is nearly invariant if $n$ is large. Hence if $\mu_n(V)$ is defined as the proportion of $S_n$ occupied by $V$, Definition 3 is satisfied.

Again, let $G$ be the group of rigid motions of the plane; let $R$ be the subgroup of rotations about 0, let $T$ be the subgroup of translations $(\xi, \eta) \to (\xi + \alpha, \eta + \beta)$, and let $T_n$ be the subset of $T$ with $\alpha^2 + \beta^2 < n^2$. Then $G = RT = TR$ represents $G$ as the topological product of the plane and the circumference of a circle. Moreover ordinary volume under this coordinatization is invariant under group translations; while the subsets $S_n = T_nR = RT_n$ are invariant under group translations by elements of $R$, and nearly so for large $n$ by elements of $T$. Hence if we define $\mu_n(V)$ as the proportion of $S_n$ occupied by $V$, Definition 3 is satisfied.

Finally, any direct product of ergodic groups is ergodic—we simply take the product of the group measures,[21]—and any homomorphic image of an ergodic group—take the image of the original measure,—is ergodic. This remark covers the extension from the cyclic to the $r$-parameter Abelian case.

On the other hand, consider groups—such as the two-parameter group with $[XY] = Y$,—whose adjoints are not unimodular. It is difficult to see how $\mu_n(T_a^{-1}VT_a)$ can ever approximate $\mu_n(V)$; hence it is probable that no such group can be "ergodic."

### 16. Digression on integration

In applying the method of F. Riesz to ergodic groups, we shall often use integrals of the form $\int_G xT_g \, d\mu_n$, and so it is necessary to give a precise meaning to such expressions.

In case $G$ is the discrete cyclic group generated by $T$, and $\mu_n(V)$ assigns to

[21] This construction has been given by Z. Lomnicki and S. Ulam, Fund. Math. 23 (1924). pp. 273–8. We omit the proof of almost invariance.

$T, T^2, \cdots, T^n$ the measure $1/n$ apiece, and to the rest of $G$ the measure zero, this is easy. In this case, the integral is simply the $n^{\text{th}}$ mean:

$$\int_G xT_g \, d\mu_n = \frac{1}{n}(xT + xT^2 + \cdots + xT^n)$$

A similar construction is possible whenever $G$ is the sum of a finite set and a set of measure zero—as with the discrete cyclic group with $n$ generators.

But in the case of continuous groups, it is necessary to use an integral for functions with values in a Banach space. Several such integrals have been defined;[22] we shall choose the definition of G. Birkhoff.

Consider the set of partitions $\pi$ of the group $G$ into finite or countable[23] measurable subsets $S_i$ of measures $\mu(S_i)$; these are transitively ordered if one defines the successors of any $\pi$ as its *subpartitions*; in fact, they have the property of Moore-Smith. Now form the set $J_\pi$ of sums

$$\sum_i \mu(S_i) x T_i \qquad (T_i \text{ in } S_i).$$

If the $J_\pi$ converge (metrically) to a limit $a$ under the ordering of partitions, then we shall say that the integral $\int_G xT_g \, d\mu$ exists and is $a$.

The only properties of this integral which we shall need are the following:

(1) If $\mu(G) = 1$, then $\int_G xT_g \, d\mu$ is in the closed convex hull $K(x)$ of the set of transforms of $x$. Hence if $a$ is a fix-point, then $\int_G aT_g \, d\mu = a$.

(2) The integral is a linear function of $x$, and the set of integrable $x$ is a subspace.

(3) If $\mu(G) = 1$, then $\left| \int_G xT_g \, d\mu \right| \leqq \sup |xT_g|$.

(4) For any linear operator $X$,

$$\left( \int_G xT_g \, d\mu \right) X = \int_G xT_g X \, d\mu.$$

(5) A change of at most $\epsilon$ in the measure function changes the integral by at most $2\epsilon \cdot \sup |xT_g|$.

These properties are not peculiar to the integral chosen; they are common

---

[22] S. Bochner, Fund. Math. 20 (1933), pp. 262–76; G. Birkhoff, Trans. Am. Math. Soc. 28 (1935), pp. 357–78; N. Dunford, ibid., pp. 441–53; B. J. Pettis, ibid. 44 (1938), pp. 277–304.

[23] We can either use finite decompositions with a finitely additive measure function and ring of measurable sets (Riemann integral), or countable decompositions with a countably additive measure function and $\sigma$-ring of measurable sets (Lebesgue integral).

to all the integrals cited. Hence any of these could be used instead of ours, without spoiling the results proved below.[24]

### 17. Equivalence of two definitions of ergodicity

We shall now prove the equivalence of Definition 2 of ergodicity, with a natural extension (to "ergodic" groups) of the usual concept of the convergence of $n^{\text{th}}$ means. This will tie up the results obtained in §§10–13 by the "minimal" method, with the theorems which we are about to obtain by a "compactness" argument.

THEOREM 7: *If $G$ is ergodic and if the integrals $s_n = \int_G xT_g\, d\mu_n$ exist, then $s_n \to a$ if and only if $x$ is "ergodic" with limit fix-point $a$, in the sense of Definition 2.*

PROOF: If $x$ is "ergodic" with limit fix-point $a$, then to any $\epsilon > 0$ corresponds a mean $T^*$ of $T_g$ making $|\, xT^* - a\,| < \epsilon$. Hence by property (3) of integrals

$$\left|\int_G (xT^*)T_g\, d\mu_n - a\right| < K\epsilon \quad \text{for all } n.$$

Again, since $T^*$ is a mean of a *finite* set of $T_g$, if $n$ is large enough, by Definition 3 *left*-multiplication by $T^*$ changes $\mu_n(V)$ by at most $\epsilon$. Hence by property (5)

$$\left|\int_G xT^*T_g\, d\mu_n - \int_G xT_g\, d\mu_n\right| < 2K\,|\,x\,|\cdot\epsilon.$$

Thus if $n$ is large enough, $|\, s_n - a\,| < K\epsilon(1 + 2\,|\,x\,|)$, whence $s_n \to a$. This proves half the theorem.

Conversely, assume $s_n \to a$. As before, for given $T^*$ and all sufficiently large $n$, $\left|\int_G xT^*T_g\, d\mu_n - s_n\right|$ is arbitrarily small—whence so is $\left|\int_G xT^*T_g\, d\mu_n - a\right|$. But by property (1) of integrals, $\int_G xT^*T_g\, d\mu_n$ is in the closed convex hull $K(xT^*)$ of the set of transforms of $xT^*$; hence we can choose a "successor" $xT^*T_1^*$ of $xT^*$ arbitrarily near $\int_G xT^*T_g\, d\mu_n$, and so arbitrarily near $a$. Conse-

---

[24] In fact, properties (1)–(4) are evident properties of any reasonable integral; only the proof of (5) offers any difficulty. To prove this under our definition, note that

$$|\sum \mu(S_i)xT_i - \sum \mu^*(S_i)xT_i\,| \leqq \sum |\,\mu(S_i)xT_i - \mu^*(S_i)xT_i\,|$$

$$\leqq \sum |\,\mu(S_i) - \mu^*(S_i)\,|\cdot|\,xT_i\,| \leqq \sup |\,xT_i\,|\cdot\sum |\,\mu(S_i) - \mu^*(S_i)\,|.$$

But if $|\,\mu(V) - \mu^*(V)\,| < \epsilon$ for all $V$, then denoting by $V$ the sum of the $S_i$ on which $\mu$ exceeds $\mu^*$, we get $\sum |\,\mu(S_i) - \mu^*(S_i)\,| = \mu(V) - \mu^*(V) < \epsilon$ for those $S_i$, and inverting the role of $\mu$ and $\mu^*$, and adding,

$$\sum |\,\mu(S_i) - \mu^*(S_i)\,| < 2\epsilon \quad \text{summed over all } S_i.$$

We now get (5) by substituting above, and passing to the limit.

quently *if* $a$ is a fix-point, then every successor of $xT^*T_1^*$ will be at most $K$ times as far from $a$—whence $x$ is "ergodic" with limit fix-point $a$.

It thus remains to show that if $\{s_n\}$ converges to a limit $a$, then $a$ is a fix-point. But indeed, for any $T$ of $G$, by the triangle inequality

$$\begin{aligned} |\, a - aT \,| &\leqq |\, a - s_n \,| + |\, s_n - s_nT \,| + |\, s_nT - aT \,| \\ &\leqq (1 + K)\,|\, s_n - a \,| + |\, s_n - s_nT \,|. \end{aligned}$$

But $|\, s_n - a \,| \to 0$ as $n \to \infty$; hence we need only show that $|\, s_n - s_nT \,| \to 0$ as well. But by property (4) of integrals,

$$|\, s_n - s_nT \,| = \left| \int_G xT_g \, d\mu_n - \int_G xT_gT \, d\mu_n \right|$$

whence by property (5) and the almost *right*-invariance of measure, this is true, completing the proof.

This last result is analogous to the Remark of §8.

## 18. Corollaries

It follows from Theorem 7 by hindsight that the convergence of the "$n^{\text{th}}$ means" $\int_G xT_g \, d\mu_n$ is really independent of the choice of an "ergodic sequence" of measure functions. For illustration, let $G$ be the cyclic group of powers $\cdots, T^{-1}, I, T, T^2, \cdots$ of a single operator $T$. One sequence of measures satisfying Definition 3 is furnished by the $n^{\text{th}}$ means $\frac{1}{n}(xT + \cdots + xT^n)$; another by the $(-n)^{\text{th}}$ means $\frac{1}{n}(xT^{-1} + \cdots + xT^{-n})$. It is a corollary of Theorem 7 that the requirement that the $n^{\text{th}}$ means converge to $a$, is equivalent to the requirement that the $(-n)^{\text{th}}$ means converge to $a$: both being equivalent to Definition 2 of ergodicity.

A more important consequence of Theorem 7 is

THEOREM 8: *If any subsequence* $\{s_{n(i)}\}$ *of the sequence* $\{s_n\}$ *of Theorem* 7 *converges even weakly, then* $x$ *is ergodic.*

PROOF: We shall form a new sequence of measure functions as follows. It is known[25] that if $s_{n(i)} \to a$ weakly, then there exists a sequence $\{t_j\}$, whose $j^{\text{th}}$ term is a mean $t_j = \sum \lambda_i^j s_{n(i)}$ of $s_{n(j)}$ and its successors, which converges metrically to $a$. But each $t_j$ is the integral

$$t_j = \sum \lambda_i^j \int_G xT_g \, d\mu_{n(i)} = \int_G xT_g \, d(\sum \lambda_i^j \mu_{n(i)})$$

[25] Banach [2], p. 134, Thm. 2. The fact that $t_j$ can be made a mean of $S_{n(j)}$ and its successors, follows as a corollary of the fact that after deleting $S_{n(1)}, \cdots, S_{n(j-1)}$, we get a sequence still converging weakly to $a$.

The above argument was introduced by F. Riesz in a form suitable for the spaces $(L_p)$; it was extended to abstract uniformly convex spaces by Kakutani; in its present form it applies to any Banach space (cf. Dunford, [7]).

of the transforms of $x$ with respect to a *mean* of the measure function $\mu_{n(j)}$ and its successors. While taking a *mean* does not increase

$$|\mu(TV) - \mu(V)| + |\mu(VT) - \mu(V)|$$

or alter the other conditions of Definition 3: the sequence $\mu_j^* = \sum \lambda_i^j \mu_{n(j)}$ of measure functions satisfies Definition 3 if $\{\mu_n\}$ does. We conclude that the means of the transforms of $x$ converge metrically to $a$, when integrated with respect to at least *one* ergodic sequence of measure functions. Hence by Theorem 7, $x$ is ergodic with limit-mean $a$.

## 19. Second main result

Theorem 8 leaves us in a position to prove a principle from which a host of special corollaries can be developed, namely

THEOREM 9: *If $G$ is ergodic, and if the transforms of $x$ lie in a "weakly compact" convex set (and are integrable), then $x$ is ergodic.*[26]

PROOF: Consider the means $s_n = \int_G xT_g \, d\mu_n$ ; they will also lie in this set. Hence we can extract a weakly convergent subsequence, and apply Theorem 8.

COROLLARY 1: *If $G$ is ergodic, and if $E$ is uniformly convex, or reflexive, or if for any other reason the unit sphere of $E$ is weakly compact, then every element of $E$ is ergodic.*

For if $E$ is uniformly convex, then it is reflexive;[27] if it is reflexive, then its unit sphere is weakly compact; if its unit sphere is weakly compact, then every bounded set is weakly compact, and so Theorem 9 applies.

But the spaces $L_p$ and $l_p$ $(p > 1)$ are uniformly convex; also, the unit sphere of any finite-dimensional Banach space is compact, hence we get more concretely,

COROLLARY 2: *If $G$ is ergodic, and $E$ is a space $L_p$ or $l_p$ $(p > 1)$, or finite-dimensional, then every element of $E$ is ergodic.*

When we come to the spaces $(L)$ and $(l)$, we know by the example of §8 that the above results no longer hold. On the other hand, it is known that if $a(x)$ and $b(x)$ are given, then the functions satisfying $a(x) \leqq f(x) \leqq b(x)$ identically, form a weakly compact convex set: any set of functions which is *lattice-theoretically* bounded, is weakly compact.

COROLLARY 3: *If the set of transforms of $x$ is bounded lattice-theoretically, if $E$ is the space $(L)$ or $(l)$, and if $G$ is ergodic, then $x$ is ergodic.*[28]

---

[26] This is a straight forward generalization of Dunford's main theorem, from the case that $G$ is an $r$-parameter Abelian group to the case that $G$ is "ergodic."

[27] For proofs of these known facts, cf. B. J. Pettis, *A proof that every uniformly convex space is reflexive*, Duke Jour. 5 (1939), pp. 249-53, and *A note on regular Banach spaces*, Bull. Am. Math. Soc. 44 (1938), pp. 420-28. This corollary is borrowed from Dunford [7], p. 641.

[28] The "weak completeness" of the space $(L)$ (Banach [2], p. 141) plus an adaptation of the argument of G. D. Birkhoff [3], made by G. Birkhoff [4], proves the weak convergence of the $n^{\text{th}}$ means of the transforms of $x$. Their strong convergence was first proved by F. Riesz [10].

Again, let $f(x)$ be any integrable function on an ergodic group $G$; the group of translations $x \to axb$ will also be ergodic. Moreover if $f(x)$ is *almost periodic*, then by definition the $f(axb)$ lie in a strongly compact convex set. Hence we conclude

COROLLARY 4: *The means of the group transforms $f(axb)$ of any almost periodic integrable function on an ergodic group, converge to a constant function.*

Actually, this is true of any almost periodic function on any group—by known results of von Neumann. But what is interesting is, that it should be demonstrable as a corollary of Riesz' ergodic theorem, for so many cases.[29]

## 20. The class of ergodic semi-groups

The examples of §13 show, among other things, that it is fruitless to hope to extend the method of F. Riesz—even using a bicompactness argument—to apply to arbitrary groups. They prove too much! We shall illustrate this, by showing that they do not permit the phenomenon of multiple fix-point limits to occur.

THEOREM 10: *If $G$ is ergodic, then the closed convex hull $K(x)$ of the set of transforms of any $x$ contains at most one fix-point.*

PROOF: Suppose $a$ and $b$ both fix-points in $K(x)$; we could then find for any $\epsilon > 0$, $T^*$ and $T_1^*$ in the convex hull $G^*$ of $G$, such that $|xT^* - a| < \epsilon$ and $|xT_1^* - b| < \epsilon$. It would follow that, for all $n$

$$\left|\int_G xT^*T_g\,d\mu_n - a\right| < K\epsilon \quad \text{and} \quad \left|\int_G xT_1^*T_g\,d\mu_n - b\right| < K\epsilon.$$

But now for sufficiently large $n$,

$$\left|\int_G xT^*T_g\,d\mu_n - \int_G xT_g\,d\mu_n\right| < 2K|x|\epsilon \quad \text{and likewise for } T_1^*.$$

Adding, we see that for any 0,

$$|a - b| < K\epsilon + 2K|x|\epsilon + 2K|x|\epsilon + K\epsilon = 2K(1 + 2|x|)\epsilon.$$

It is now obvious that $a = b$, completing the proof.

HARVARD UNIVERSITY AND
INSTITUTE FOR ADVANCED STUDY

### BIBLIOGRAPHY

[1] L. ALAOGLU and G. BIRKHOFF, *General ergodic theorems*, Proc. Nat. Acad. Sci. 25 (1939). pp. 628–30.
[2] S. BANACH, *Théorie des operations linéaires*, Warsaw, 1932.
[3] G. D. BIRKHOFF, *Proof of a recurrence theorem for strongly transitive systems, and proof of the ergodic theorem*, Proc. Nat. Acad. Sci. 17 (1931), pp. 650–60.

[29] It follows from a remark of A. Weil, *Sur les fonctions presque-periodiques de M. von Neumann*, Comptes Rendus 200 (1935), p. 38, that it includes all continuous a.p. functions on Lie groups.

[4] G. Birkhoff, *Dependent probabilities and spaces (L)*, ibid. 24 (1938), pp. 154–9.
[5] G. Birkhoff, *The mean ergodic theorem*, Duke Jour. 5 (1939), pp. 635–46.
[6] G. Birkhoff, *An ergodic theorem for general semi-groups*, Proc. Nat. Acad. Sci. 25 (1939), pp. 625–7.
[7] N. Dunford, *A mean ergodic theorem*, Duke Jour. 5 (1939), pp. 635–46.
[8] S. Kakutani, *Weak convergence in uniformly convex spaces*, Tohoku Math. Jour. 45 (1939), pp. 188–93.
[9] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. 18 (1932), pp. 70–82.
[10] F. Riesz, *Some mean ergodic theorems*, Jour. Lond. Math. Soc. 13 (1938), pp. 274–8.
[11] N. Wiener, *The ergodic theorem*, Duke Jour. 5 (1939), pp. 1–18.
[12] K. Yosida, *Mean ergodic theorem in Banach spaces*, Proc. Imp. Acad. Tokyo 14 (1938), pp. 292–4.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# ON EXTREMAL PROPERTIES OF THE DERIVATIVES OF POLYNOMIALS

BY P. ERDÖS

(Received March 2, 1939)

Throughout this paper let $f(x)$ be a polynomial of degree $n$ satisfying the inequality $|f(x)| \leqq 1$ for $-1 \leqq x \leqq 1$. A. Markoff[1] showed that for $-1 \leqq x \leqq 1$, $|f'(x)| \leqq n^2$. Equality is obtained only for the Tchebicheff polynomial $T_n(x)$ for $x = \pm 1$. In the present paper we shall prove the following analogous

THEOREM.[2] *Suppose $f(x)$ has only real roots and no roots in $-1, +1$; then for $-1 \leqq x \leqq 1$, $|f'(x)| < \frac{1}{2}en$. This is the best possible result.*

PROOF. We distinguish two cases. First we assume that $f'(x)$ has a root $x_0$ such that $-1 \leqq x_0 \leqq 1$. A simple linear transformation enables us to put $\max f(x) = 1$ for $-1 \leqq x \leqq 1$ and $f(-1) = f(+1) = 0$. We prove the following

LEMMA. *Suppose $f(x)$ has only real roots none of which lie in the interval $-1, +1$ and let $f(-1) = f(+1) = 0$, $\max_{-1 \leqq x \leqq 1} f(x) = 1$, then*

$$f(x) < e\,\frac{1+x}{1+x_0}.$$

PROOF. Put $x_0 - x = d$. Without loss of generality we may assume $-1 < x < x_0$. Then if $x_i$ $(i = 1, 2, \cdots, n)$, $x_1 = -1$, $x_2 = +1$ denote the roots of $f(x)$, we have

$$(1) \qquad 1 = f(x_0) = c \prod_{i \leqq n} (x_0 - x_i), \quad \text{and} \quad \sum_{i \leqq n} \frac{1}{x_0 - x_i} = 0 \left(f(x) = cx^n + \cdots\right).$$

Now

$$f(x) = c \prod_{i \leqq n} (x - x_i) = c(1+x) \prod_{i=2}^{n} (-d + x_0 - x_i).$$

Thus

$$f(x) = \frac{f(x)}{f(x_0)} = \frac{1+x}{1+x_0} \prod_{i=2}^{n} \left(1 - \frac{d}{x_0 - x_i}\right).$$

But by (1)

$$\sum_{i=2}^{n} \frac{1}{x_0 - x_i} = -\frac{1}{1+x_0}.$$

Hence from $\prod (1 + a_j) < \exp \sum a_j$, we have

$$f(x) < \frac{1+x}{1+x_0} \exp \frac{d}{1+x_0} < e\,\frac{1+x}{1+x_0}. \quad \text{q.e.d.}$$

[1] A. Markoff, Abh. Akad. Wiss. St. Petersburg, 1889, vol. 62, pp. 1–24.
[2] The same result was obtained by Mr. Erőd by a different method.

By a slightly longer calculation we could show that

$$f(x) \leqq \left(1 - \frac{1}{n}\right)^{-n+1} \frac{1+x}{1+x_0}, \tag{2}$$

equality occurring only for

$$f(x) = \frac{n}{2^n\left(1 - \frac{1}{n}\right)^{n-1}} (x+1)(x-1)^{n-1}$$

and

$$f(x) = \frac{n}{2^n\left(1 - \frac{1}{n}\right)^{n-1}} (x+1)^{n-1}(x-1).$$

In these cases

$$f'(1) = \frac{n}{2\left(1 - \frac{1}{n}\right)^{n-1}} \to \frac{e}{2} n.$$

Thus it can be shown by an easy calculation that the constant $e$ of our Lemma cannot be improved. We have

$$\sum_{x_i > x_0} \frac{1}{x_i - x_0} = \sum_{x_i < x_0} \frac{1}{x_0 - x_i} \leqq \min\left(\frac{k}{1-x_0}, \frac{n-k}{1+x_0}\right) \leqq \frac{n}{2}, \tag{3}$$

where $k$ denotes the number of roots $> x_0$ of $f(x)$. Now by our Lemma we obtain

$$f(x) = f(x) \sum_{i=1}^{n} \frac{1}{x - x_i} < e \frac{1+x}{1+x_0} \sum_{x_i < x_0} \frac{1}{x - x_i} \leqq e \frac{1+x}{1+x_0} \frac{1+x_0}{1+x} \sum_{x_i < x_0} \frac{1}{x_0 - x_i}$$

$$< \frac{e}{2} n,$$

by (3) which proves our Theorem for the first case. From (2) we can deduce that

$$|f'(x)| \leqq \frac{n}{2\left(1 - \frac{1}{n}\right)^{n-1}} \quad \text{for} \quad -1 \leqq x \leqq 1.$$

Suppose now that $f'(x) \neq 0$ for $-1 \leqq x \leqq 1$. A linear transformation enables us to put $f(-1) = 0$, $f(+1) = 1$. The roots of $f(x)$ are $x_1 = -1$, $x_2$, $x_3, \cdots, x_n$. Now

$$f'(1) \leqq \sum_{x_i < 1} \frac{1}{1 - x_i} \leqq \frac{n}{2}.$$

We have as in our Lemma

$$f(x) = \frac{f(x)}{f(1)} = \frac{1+x}{2}\prod_{i=2}^{n}\left(\frac{x-x_i}{1-x_i}\right) = \frac{1+x}{2}\prod_{i=2}^{n}\left(1-\frac{1-x}{1-x_i}\right)$$
$$< \frac{1+x}{2}\exp\frac{1-x}{2} < e\frac{1+x}{2}.$$

In the above we used the fact that

$$\frac{1}{2} + \sum_{i=2}^{n}\frac{1}{1-x_i} \geqq 0,$$

along with the inequality $\prod(1+a_j) < \exp\sum a_j$. Thus

$$f'(x) = f(x)\sum_{i=1}^{n}\frac{1}{x-x_i} < e\frac{1+x}{2}\sum_{x_i<x}\frac{1}{x-x_i} \leqq e\frac{1+x}{2}\frac{2}{1+x}\sum_{x_i<1}\frac{1}{1-x_i} \leqq \frac{e}{2}n.$$

This completes the proof of our Theorem. A slightly longer calculation would show that in the second case $f'(x) \leqq \frac{1}{2}n$, where equality holds only for $f(x) = (1 \pm x)^n/2^n$. A theorem of S. Bernstein[3] states that for $-1 \leqq x \leqq 1$, $|f'(x)| \leqq n/(1-x^2)^{\frac{1}{2}}$. For every subinterval of $-1$, $+1$ this result is very much better than the theorem of Markoff. By analogy we prove

THEOREM.[4] *Let $f(x)$ be a (real valued) polynomial having no root in the interior of the unit circle; then for $-1+c < x < 1+c$, $|f'(x)| < \frac{4}{c^2}n^{\frac{1}{2}}$ for $n > n_0$.*

PROOF. Suppose that for a certain $x_0$ in $-1+c$, $1-c$ $|f'(x_0)| \geqq \frac{4}{c^2}$. Put $|x - x_0| < n^{-\frac{1}{2}}$ and denote by $x_1, x_2, \cdots, x_n$ the roots of $f(x)$. Then since $f(x)$ has no root in the unit circle

$$\left|\frac{1}{x-x_i} - \frac{1}{x_0-x_i}\right| \leqq \frac{1}{c} - 1/c - n^{-\frac{1}{2}} = n^{-\frac{1}{2}}/c(c-n^{-\frac{1}{2}}) < 2/c^2n^{\frac{1}{2}} \quad \text{for} \quad n > n_0.$$

Without loss of generality we may assume that $f'(x_0) > 0$. Then

$$f'(x) = f(x)\sum_{i=1}^{n}\frac{1}{x-x_i} > f(x)\sum_{i=1}^{n}\frac{1}{x_0-x_i} - \frac{2n^{\frac{1}{2}}}{c^2} > 0. \tag{4}$$

since

$$\sum_{i=1}^{n}\frac{1}{x_0-x_i} = \frac{f'(x_0)}{f(x_0)} \geqq \frac{4}{c^2}n^{\frac{1}{2}}. \tag{5}$$

Thus $f(x)$ increases for $x_0 < x < x_0 + n^{-\frac{1}{2}}$. Hence by (4) and (5) $f'(x) > \frac{f'(x_0)}{2} > \frac{2n^{\frac{1}{2}}}{c^2}$. But

$$1 > \int_{x_0}^{x_0+n^{-}} f'(x)\,dx > \frac{2}{c^2}$$

[3] S. Bernstein, Belg. Mém. 1912, p. 19.

[4] This problem was suggested to me by Professor D. R. Curtiss.

which leads to a contradiction. Put

$$f(x) = \frac{1}{e}(x^2 - 1)^n(1 + x)^{[n^{\frac{1}{2}}]}{}^5$$

Writing $x = an^{-\frac{1}{2}}$ we have,

$$|f(x)| = \frac{1}{e}\left(1 - \frac{a^2}{n}\right)^n (1 + an^{-\frac{1}{2}})^{[n^{\frac{1}{2}}]} < \frac{1}{e}e^{-a^2+a} < 1.$$

But $|f'(0)| = \dfrac{[n^{\frac{1}{2}}]}{e}$ which shows that in Theorem 2. $n^{\frac{1}{2}}$ cannot be replaced by any function tending to infinity more slowly.

INSTITUTE FOR ADVANCED STUDY.

[5] $[n^{\frac{1}{2}}]$ denotes the greatest integer not exceeding $n^{\frac{1}{2}}$.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# AN ESTIMATE OF THE WEIERSTRASS E-FUNCTION

BY E. J. MC SHANE

(Received April 22, 1939)

In a recent paper[1] W. T. Reid presented a highly interesting proof of a sufficiency theorem for the general Bolza problem, using expansion methods. One of the several inequalities on which the proof rests is an estimate for the values of the E-function when the strengthened Weierstrass condition is assumed to hold. In Reid's proof of this inequality, and also in another proof which I devised independently, it was necessary to assume that the curve is non-singular. The analogous theorem for free problems needs no such assumption, so that it would seem to be of some interest to establish this inequality on the E-function without assuming non-singularity. Here such a proof is presented. The lemma on which the proof rests seems to have some intrinsic interest, in that it shows that the canonical variables $(x, y, z)$ are in one-to-one continuous correspondence with the admissible sets $(x, y, y', \lambda)$ if the strengthened Weierstrass condition holds, even though the curve is singular.

## 1. Notation and Definitions

We shall assume that the functions $f(x, y, r)$ and $\varphi^\alpha(x, y, r)$, $\alpha = 1, \cdots, m < n$ are defined and continuous for all points $(x, y, r) \equiv (x, y^1, \cdots, y^n, r^1, \cdots, r^n)$ in a set $\mathfrak{R}$, and have partial derivatives with respect to the $r^i$ which are defined and continuous on $\mathfrak{R}$. These are the only partial derivatives which will appear in this note, so we save complicated subscripts by agreeing that the components of a vector, such as $y$ or $r$, shall be indicated by superscripts, while a subscript $i$ will indicate the partial derivative with respect to $r^i$. Also, we use the tensor summation convention, so that for example

$$r^i \varphi_i^\alpha = \sum_{i=1}^{n} r^i \frac{\partial \varphi^\alpha}{\partial r^i}.$$

The length of a vector will be denoted by vertical bars $| \cdot |$; thus $| r | = (r^i r^i)^{\frac{1}{2}}$.

A set $(x, y, r)$ is *admissible* if it is in $\mathfrak{R}$ and satisfies the equations $\varphi^\alpha(x, y, r) = 0$. For convenience, we shall use the words "the set $(x, y, r, \lambda)$ is admissible" as

[1] W. T. Reid, *Sufficient conditions by expansion methods for the problem of Bolza in the calculus of variations*, Annals of Mathematics, 38 (1937) pp. 662–678. See also W. T. Reid, *A direct expansion proof of sufficient conditions for the non-parametric problem of Bolza*, Transactions of the American Mathematical Society 42 (1937), pp. 183–190.

an abbreviation for "the set $(x, y, r, \lambda)$ is such that $(x, y, r)$ is admissible." As usual, we define

$$F(x, y, r, \lambda) \equiv \lambda^0 f(x, y, r) + \lambda^a \varphi^a(x, y, r),$$

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) \equiv F(x, y, \bar{r}, \lambda) - F(x, y, r, \lambda) - (\bar{r}^i - r^i) F_i(x, y, r, \lambda).$$

## 2. Statement of the Theorem and of the lemma

Before stating the theorem whose proof is the principal object of this note, we wish to make a comment on the multiplier $\lambda^0$. In normal problems and certain anormal problems, this can be taken to be 1. In anormal problems it can be taken to be 0. In any case, after its introduction in hypothesis (2) of the theorem below it is supposed to be fixed, and not thereafter referred to. Thus, for example, when we speak of $(x, y, r, \lambda)$-space we understand $\lambda$ to be $(\lambda', \cdots, \lambda^m)$, so that $(x, y, r, \lambda)$-space is $(2n + m + 1)$-dimensional.

Now we state our theorem.

THEOREM. *Let the following hypotheses be satisfied.*

(1) *The curve* $C: y = \gamma(x)$, $x_1 \leqq x \leqq x_2$ *is such that* $\gamma'(x)$ *is continuous and* $(x, \gamma, \gamma')$ *admissible for* $x_1 \leqq x \leqq x_2$.

(2) *There are multipliers* $\lambda^0$, $\lambda^1(x), \cdots, \lambda^m(x)$, *where* $\lambda^0$ *is constant and the* $\lambda^i(x)$ *are continuous, for which condition* $II'_N$[2] *is satisfied on* $C$; *that is, in* $(x, y, r, \lambda)$*-space there is a neighborhood* $N$ *of the point set* $[x_1 \leqq x \leqq x_2, \gamma(x), \gamma'(x), \lambda(x)]$ *such that*

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) > 0$$

*whenever* $(x, y, r, \lambda)$ *and* $(x, y, \bar{r}, \lambda)$ *are distinct admissible sets and* $(x, y, r, \lambda)$ *is in* $N$.

(3) *The matrix* $\| \varphi_i^a(x, \gamma, \gamma') \|$ *has rank* $m$ *on the interval* $x_1 \leqq x \leqq x_2$.

*Then for every positive number* $\sigma$ *there is a positive number* $\mu$ *and a neighborhood* $N_0$ *of the points* $(x, \gamma(x), \gamma'(x), \lambda(x))$ *in* $(x, y, r, \lambda)$*-space such that the inequality*

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) \geqq \mu \, | \, \bar{r} - r \, | \tag{2.1}$$

*holds whenever* $(x, y, r, \lambda)$ *is in* $N_0$, *the sets* $(x, y, r)$ *and* $(x, y, \bar{r})$ *are admissible, and* $| \, \bar{r} - r \, | \geqq \sigma$.

In proving this theorem we shall make use of the "canonical variables" $(x, y, z)$. For every admissible set $(x, y, r, \lambda)$ the corresponding $(x, y, z)$ is defined by the equations $x = x$, $y = y$, and

$$z^i = z^i(x, y, r, \lambda) = F_i(x, y, r, \lambda), \qquad i = 1, \cdots, n. \tag{2.2}$$

In particular, the elements of $C$ are transformed into $(x, \gamma(x), z(x))$, where

$$z^i(x) = F_i(x, \gamma(x), \gamma'(x), \lambda^0, \lambda(x)), \qquad i = 1, \cdots, n, \; x_1 \leqq x \leqq x_2 \tag{2.3}$$

[2] Cf. Reid, l.c., p. 664.

These variables are much used in studying Lagrange and Bolza problems. For "non-singular" curves $C$ it is well known[3] that equations (2.2) have solutions $r(x, y, z)$ and $\lambda(x, y, z)$ which are single-valued, admissible and continuously differentiable near the sets $(x, \gamma(x), z(x))$. Here we are not assuming non-singularity, and it is interesting in itself, as well as useful in the following pages, to see how much of this solvability persists when for non-singularity we substitute the condition $\Pi'_N$. The result is contained in the following lemma.[4]

LEMMA: *Under the hypotheses of the theorem, there is a neighborhood $N_1$ of the sets $(x, \gamma(x), \gamma(x), \lambda(x))$ such that all admissible sets in $N_1$ are mapped topologically on a neighborhood $Z$ in $(x, y, z)$-space of the point-set $(x, \gamma(x), z(x))$, $x_1 \leqq x \leqq x_2$.*

Thus the unique solvability of equations (2.2) and the continuity of the solutions persist, but the solutions are not necessarily differentiable.

### 3. Proof of the Lemma

Bliss has shown[5] that if hypothesis (3) is satisfied, it is possible to find $n - m$ analytic functions $\varphi^\beta(x, y, r)$, $\beta = m + 1, \cdots, n$, such that the square matrix

$$\| \varphi^i_j(x, \gamma, \gamma') \|, \qquad i, j = 1, \cdots, n \tag{3.1}$$

is non-singular for $x_1 \leqq x \leqq x_2$. We may suppose that the neighborhood $N$ of hypothesis (2) is small enough so that

$$\| \varphi^i_j(x, y, r) \| \tag{3.2}$$

is non-singular whenever $(x, y, r, \lambda)$ is in $N$. By adding to $\varphi^\beta(x, y, r)$ the function $-\varphi^\beta(x, \gamma(x), \gamma'(x))$ we obtain the equations

$$\varphi^i(x, \gamma, \gamma') = 0, \quad i = 1, \cdots, n, x_1 \leqq x \leqq x_2, \tag{3.3}$$

the functions $\varphi^\beta$ still being continuously differentiable and statement (3.2) remaining valid.

Consider now the equations

$$\begin{aligned} \varphi^\alpha(x, y, r) - u^\alpha &= 0, & \alpha &= 1, \cdots, m, \\ \varphi^\beta(x, y, r) - v^\beta &= 0, & \beta &= m + 1, \cdots, n. \end{aligned} \tag{3.4}$$

These equations have the initial solution $y = \gamma(x)$, $r = \gamma'(x)$, $u = v = 0$, $x_1 \leqq x \leqq x_2$, and the matrix of the partial derivatives of the left members with respect to the $r^i$ is $\| \varphi^i_j(x, y, r) \|$, which is non-singular on the initial

---

[3] See, e.g. Bliss, *The problem of Lagrange in the calculus of variations*, American Journal of Mathematics, vol. 52 (1930); in particular, pp. 685–687.

[4] The conclusion of the lemma remains valid, as the proof will show, if we weaken hypothesis (2) by assuming only that $\mathcal{E}(x, y, r, \bar{r}, \lambda) > 0$ whenever $(x, y, r, \lambda)$ and $(x, y, \bar{r}, \lambda)$ are distinct admissible sets *both* in $N$.

[5] G. A. Bliss, *The problem of Mayer with variable end-points*, Transactions of the American Mathematical Society, vol. 19 (1918); in particular, pp. 307, 312.

solution. Hence these equations can be solved for the $r^i$. That is, there exists a set of functions

$$r^i = \rho^i(x, y, u, v), \tag{3.5}$$

defined and continuous on a neighborhood $U$ in $(x, y, u, v)$-space of the point-set $[x_1 \leqq x \leqq x_2, y = \gamma(x), u = 0, v = 0]$, with which equations (3.4) hold identically. Also, these satisfy the initial conditions:

$$\rho^i(x, \gamma(x), 0, 0) = \gamma^{i'}(x).$$

We have thus established the preliminary statement

(A) *The mappings* (3.4) *and* (3.5) *are inverse, and map a neighborhood* $U$ *of* $[x, \gamma(x), 0, 0]$ *topologically on a neighborhood* $N_0$ *of the points* $[x, \gamma(x), \gamma'(x)]$.

In particular, the point $(x, y, r)$ of $N_0$ is admissible if and only if its image $(x, y, u, v)$ in $U$ has $u = 0$, by equations (3.4). So if we define $r^i(x, y, v) \equiv \rho^i(x, y, 0, v)$, we find that

(B) *The mappings*

$$v^\beta = \varphi^\beta(x, y, r) \tag{3.6}$$

*and*

$$r^i = r^i(x, y, v) \tag{3.7}$$

*are inverse, and map a neighborhood* $U_0$ *of the points*

$$[x_1 \leqq x \leqq x_2, y = \gamma(x), v = 0]$$

*topologically on the set of all admissible elements* $(x, y, r)$ *in* $N_0$.

As an immediate consequence of the continuity of the mappings, we have

(C) *If the neighborhood* $U_0$ *and the positive number* $\epsilon$ *are small enough, then for every* $(x, y, v, \lambda)$ *in the closure of the set*

$$V_0: \quad (x, y, v) \text{ in } U_0, \qquad |\lambda - \lambda(x)| < \epsilon$$

*the point* $(x, y, r(x, y, v), \lambda)$ *is in the set* $N$ *occurring in the statement of property* $II'_N$; *and therefore if* $(x, y, v, \lambda)$ *and* $(x, y, \bar{v}, \lambda)$ *are both in the closure* $\bar{V}_0$ *of* $V_0$, *the inequality*

$$\mathcal{E}(x, y, r(x, y, v), r(x, y, \bar{v}), \lambda) > 0 \tag{3.8}$$

*is satisfied unless* $r(x, y, v) = r(x, y, \bar{v})$.

Next we map the closure $\bar{V}_0$ of $V_0$ on a set of points $(x, y, z)$ by the equations $x = x$, $y = y$ and

$$z^i = F_i(x, y, r(x, y, v), \lambda) = \lambda^0 f_i(x, y, r(x, y, v)) + \lambda^\alpha \varphi_i^\alpha(x, y, r(x, y, v)). \tag{3.9}$$

This mapping is obviously continuous; we shall now prove it one-to-one. Suppose that two points $(x, y, v, \lambda)$ and $(\bar{x}, \bar{y}, \bar{v}, \bar{\lambda})$ of $\bar{V}_0$ have the same image under (3.9), so that $\bar{x} = x$, $\bar{y} = y$ and

$$z^i(x, y, v, \lambda) = z^i(\bar{x}, \bar{y}, \bar{v}, \bar{\lambda}). \tag{3.10}$$

For brevity we denote $r(x, y, v)$ and $r(\bar{x}, \bar{y}, \bar{v})$ by $r$ and $\bar{r}$ respectively. By equations (3.9) and the definition of the $\mathfrak{E}$ function,

$$\mathfrak{E}(x, y, r, \bar{r}, \lambda) = \lambda^0 f(x, y, \bar{r}) - \lambda^0 f(x, y, r) - (\bar{r}^i - r^i) z^i(x, y, v, \lambda),$$

$$\mathfrak{E}(x, y, \bar{r}, r, \bar{\lambda}) = \lambda^0 f(x, y, r) - \lambda^0 f(x, y, \bar{r}) - (r^i - \bar{r}^i) z^i(x, y, \bar{v}, \bar{\lambda}).$$

Equations (3.10) show that these are negatives of each other. But by statement (C) both would be positive if $\bar{r} \neq r$. Hence we conclude $\bar{r} = r$. Now in equation (3.10) we substitute the definition of $z^i$, recalling that $(\bar{x}, \bar{y}, \bar{r}) = (x, y, r)$. We obtain

$$(\bar{\lambda}^\alpha - \lambda^\alpha)\varphi_i^\alpha(x, y, r) = 0.$$

But this, with hypothesis (3), implies $\bar{\lambda} = \lambda$. We have thus shown that equations (3.10) imply that $(\bar{x}, \bar{y}, \bar{v}, \bar{\lambda}) = (x, y, v, \lambda)$, and the mapping is one-to-one.

Since a mapping which is continuous and one-to-one on a bounded closed set has a continuous inverse, the mapping (3.9) is topological. In particular, $V_0$ is mapped topologically on a set $Z$, which is then necessarily open.

Now let $N_1$ be the set of points $(x, y, r, \lambda)$ with $(x, y, r)$ in $N_0$ (cf. statement (B)) and $|\lambda - \lambda(x)| < \epsilon$ (cf. statement C). The set of admissible elements $(x, y, r, \lambda)$ in $N_1$ is mapped topologically by (3.6) on $V_0$. This is in turn mapped topologically on $Z$ by equations (3.9). Hence $Z$ is the topological image of the set of admissible elements in $N_1$. Since $Z$ is open and contains the point-set $(x, \gamma(x), z(x))$, it is a neighborhood of that point-set, and the lemma is established.

## 4. Proof of the Theorem

For an arbitrary positive number $\delta$, let $Z(\delta)$ be the set of points $(x, y, z)$ such that

$$x_1 - \delta < x < x_2 + \delta, \qquad |y - \gamma(x)| < \delta, \qquad |z - z(x)| < \delta.$$

If $\delta$ is sufficiently small, $Z(\delta)$ is contained in the set $Z$ of the lemma.

Now let $\sigma$ be an arbitrary positive number. We choose a positive $\mu$ so small that (i) the closure of $Z(3\mu)$ is in the set $Z$ of the lemma, and (ii) its image $N_3$ under the mapping of $Z$ on $N_1$ is so small that for every two points $(x, y, r_1, \lambda)$ and $(x, y, r_2, \lambda)$ of $N_3$ the inequality

$$|r_1 - r_2| < \sigma/2 \tag{4.1}$$

holds. We define $N_0$ to be the image of $Z(\mu)$ under the mapping of $Z$ on $N_1$.

Let $(x, y, r, \lambda)$ be in $N_0$ and (consequently) admissible, and let $(x, y, \bar{r}, \lambda)$ be admissible and have

$$|\bar{r} - r| \geqq \sigma. \tag{4.2}$$

The image of $(x, y, r, \lambda)$ in $Z(\mu)$ we denote by $(x, y, z)$. Considering $(x, y)$ fixed, from $(x, y, z)$ we construct a half-line in the direction from $r$ to $\bar{r}$. On

this half-line there is a first point $z_0$ not in $Z(3\mu)$, hence on the boundary of $Z(3\mu)$. Then

$$z_0^i - z^i = k(\bar{r}^i - r^i), \qquad i = 1, \cdots, n, \ k > 0. \tag{4.3}$$

Since $(x, y, z)$ is in $Z(\mu)$ we obviously have

$$| z_0 - z | > 2\mu. \tag{4.4}$$

From the definition of the $\mathcal{E}$-function we compute

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) = \mathcal{E}(x, y, r_0, \bar{r}, \lambda) + \mathcal{E}(x, y, r, r_0, \lambda) + (\bar{r}^i - r_0^i)(z_0^i - z^i). \tag{4.5}$$

By hypothesis (2), the first two terms on the right are non-negative, so

$$\begin{aligned} \mathcal{E}(x, y, r, \bar{r}, \lambda) &\geqq (\bar{r}^i - r_0^i)(z_0^i - z^i) \\ &= (\bar{r}^i - r^i)(z_0^i - z^i) + (r^i - r_0^i)(z_0^i - z^i) \\ &= | \bar{r} - r | \cdot | z_0 - z | + (r^i - r_0^i)(z_0^i - z^i) \\ &\geqq | z_0 - z | \, (| \bar{r} - r | - | r - r_0 |). \end{aligned} \tag{4.6}$$

Now $| \bar{r} - r | \geqq \sigma > 2 | r - r_0 |$, so, recalling (4.4),

$$\begin{aligned} \mathcal{E}(x, y, r, \bar{r}, \lambda) &\geqq \tfrac{1}{2} | z_0 - z | \cdot | \bar{r} - r | \\ &\geqq \mu | \bar{r} - r |. \end{aligned}$$

The proof of the theorem is complete.

5. An investigation of the above proof shows that we have not used the full strength of hypothesis (2). In fact we can state the following slight generalization of our theorem.

COROLLARY 1. *In the theorem, hypothesis* (2) *can be replaced by the following hypothesis.*

(2′) *There are multipliers* $\lambda^0, \lambda^1(x), \cdots, \lambda^m(x)$ *(where* $\lambda^0$ *is constant and the* $\lambda^i(x)$ *are continuous) and a neighborhood* $N$ *of the point-set* $[x_1 \leqq x \leqq x_2, \gamma(x), \gamma'(x), \lambda(x)]$ *in* $(x, y, r, \lambda)$*-space such that*

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) > 0 \tag{i}$$

*whenever* $(x, y, r, \lambda)$ *and* $(x, y, \bar{r}, \lambda)$ *are distinct admissible elements of* $N$, *and*

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) \geqq 0 \tag{ii}$$

*whenever* $(x, y, r, \lambda)$ *and* $(x, y, \bar{r}, \lambda)$ *are both admissible and* $(x, y, r, \lambda)$ *is in* $N$.

For in proving the lemma we used only elements $(x, y, r, \lambda)$ in $N$, and in deducing (4.6) from (4.5) we needed only to know that the first two terms in (4.5) are non-negative.

COROLLARY 2. *Under the hypotheses of Corollary* 1, *condition* $II'_N$ *is satisfied.*

We may suppose that the neighborhood $N_0$ of the conclusion of the theorem is contained in $N$. Let $N_4$ be a neighborhood of the set $[x_1 \leqq x \leqq x_2, \gamma(x),$

$\gamma'(x)$, $\lambda(x)$] whose closure $\bar{N}_4$ is in $N_0$. Then $\bar{N}_4$ has a positive distance from the complement of $N_0$; we denote this distance by $\sigma$. Now let $(x, y, r, \lambda)$ be an admissible element of $N_4$, and let $(x, y, \bar{r}, \lambda)$ be admissible and distinct from $(x, y, r, \lambda)$. The inequality

$$\mathcal{E}(x, y, r, \bar{r}, \lambda) > 0$$

holds by the conclusion of the theorem if $(x, y, \bar{r}, \lambda)$ is not in $N_0$, and by hypothesis $(2', \text{i})$ if it is in $N_0$.

Corollary 2 is a generalization of a recent theorem of Hestenes and Reid.[6]

THE UNIVERSITY OF VIRGINIA.

[6] Abstracted in Bull. Amer. Math. Soc., vol. 45, No. 1, (1939), p. 72. [Added in proof: Now published in Bull. Amer. Math. Soc., vol. 45 (1939), pp. 471–473.]

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# ON A CONTINUUM OF THE POWER $2^{\aleph_i}$

BY ISAIAH MAXIMOFF

(Received March 16, 1939)

Denote by $\Omega_i$ the first transfinite number which is preceded by $\aleph_i$ transfinite numbers. Let

$$x = \{x_0, x_1, x_2, \cdots, x_\alpha, \cdots, x_\beta\} \qquad (0 \leqq \beta < \Omega_i) \tag{1}$$

be a sequence of transfinite numbers $x_\alpha$, $1 \leqq x_\alpha < \Omega_i$. This sequence will be called a *rational* point of a transfinite space $E_x^{(i)}$ of order $i$. Denote by $R_x^{(i)}$ the set of all rational points of the space $E_x^{(i)}$.

## The Set $R_x^{(i)}$

DEFINITION 1. *We shall say that the rational point*

$$x' = \{x_0', x_1', x_2', \cdots, x_{\alpha'}'\} \text{ is equal to the point}$$
$$x'' = \{x_0'', x_1'', x_2'', \cdots, x_{\alpha''}''\} \text{ in the following cases:}$$

(i) *either* $\alpha' = \alpha''$ *and then* $x_0' = x_0''$, $x_1' = x_1''$, $\cdots$, $x_{\alpha'}' = x_{\alpha''}''$;
(ii) *or* $\alpha' = \alpha'' + 1$ *and then* $x_0' = x_0''$, $x_1' = x_1''$, $\cdots$, $x_{\alpha''}' + 1 = x_{\alpha''}''$, $x_{\alpha'}' = 1$;
(iii) *or* $\alpha' + 1 = \alpha''$ *and then* $x_0' = x_0''$, $x_1' = x_1''$, $\cdots$, $x_{\alpha'}' = x_{\alpha'}'' + 1, x_{\alpha''}'' = 1$.

In order to express that a point $p$ lies to the left of a point $q$ we shall write: $p < q$, or else $q > p$. Let $x'$ and $x''$ be rational points. We consider the following cases:

THE FIRST CASE: $x_0' \neq x_0''$. In this case we make the following

DEFINITION 2. *We shall say that the point $x'$ lies to the left of the point $x''$ if $x_0' < x_0''$.*

In order to treat the case $x_0' = x_0''$, we make the following definitions

DEFINITION 3. *A number $\beta$ is said to be even if it may be written in the form $\beta = \beta' + \beta''$ where $\beta'$ is a number of the second species*[1] *and $\beta''$ is either an even natural number or zero.*

DEFINITION 4. *If, on the contrary, $\alpha$ is not an even number or even, it is called an odd number.*

Returning to the numbers $x' = \{x_0', x_1', x_2', \cdots, x_{\alpha'}'\}$, $x'' = \{x_0'', x_1'', x_2'', \cdots, x_{\alpha''}''\}$ we denote by $\alpha$ the greater of the numbers $\alpha'$, $\alpha''$ and suppose that $x_n' \neq x_n''$, but $x_{n_1}' = x_{n_1}''$ for all $n_1 < n$ where $n \leqq \alpha$.

[1] Baire. Leçons sur les fonctions discontinues, Paris, 1930, p. 39.

DEFINITION 5. *Now we shall say that the point $x'$ lies to the left of the point $x''$ in the following cases*:

(i) *either $n$ is an even number and then* $\{x'_n, x'_{n+1}, \cdots, x'_{\alpha'}\} < \{x''_n, x''_{n+1}, \cdots, x''_{\alpha''}\}$;

(ii) *or $n$ is an odd number and then* $\{x'_n, x''_{n+1}, \cdots, x'_{\alpha'}\} > \{x''_n, x''_{n+1}, \cdots, x''_{\alpha''}\}$.

We may now enunciate the following

THEOREM I. *$R_x^{(i)}$ is an ordered set.*

This theorem is a corollary of the preceding definitions.

THEOREM II. *Let $x'$ and $x''$ be any rational points such that $x' < x''$, then there exists a rational point $x$ such that $x' < x < x''$.*

PROOF. We consider the following particular cases.

THE FIRST CASE: $\alpha' = 0$, $\alpha'' = 0$, $x'_0 < x''_0$. In this case the point $x = \{x'_0, 2\}$ lies between $x'$ and $x''$. If $x' = \{x'_0, 2\}$, $x'' = \{x'_0, 1\}$ we have $x' < \{x'_0, 1, 3\} < x''$.

THE SECOND CASE: $\alpha' = 0$, $\alpha'' \geqq 1$, $x'_0 < x''_0$. In this case the point $x = \{x''_0\}$ lies between $x'$ and $x''$ .

THE THIRD CASE: $\alpha' = 1$, $\alpha'' = 0$, $x'_0 < x''_0$, $x'_1 = 1$. Since $\{x'_0, 1\} = \{x'_0 + 1\} = x'$ and $x' < x''$, we have $x'_0 + 1 < x''_0$. In this case the point $x = \{(x_0 + 1), 2\}$ lies between $x'$ and $x''$.

THE FOURTH CASE: $\alpha' = 1$, $\alpha'' = 0$, $x'_0 < x''_0$, $x'_1 = 2$. In this case $x = \{x_0, 1, 2\}$ lies between $x' = \{x'_0, 1\}$ and $x'' = \{x''_0\}$.

THE FIFTH CASE: $\alpha' = 1$, $\alpha'' = 0$, $x'_0 < x''_0$, $x'_1 \geqq 3$. In this case the point $x = \{x'_0, 2\}$ lies between $x'$ and $x''$.

THE SIXTH CASE: $\alpha' \geqq 2$, $\alpha'' = 0$, $x'_0 < x''_0$. In this case the point $x = \{x'_0, x'_1, (x'_2 + 1)\}$ lies between $x' = \{x'_0, x'_1, x'_2\}$ and $x'' = \{x''_0\}$.

THE SEVENTH CASE: $\alpha' = \alpha'' = 1$, $x'_0 < x''_0$. If one of the numbers $x'_1$ and $x''_1$ is equal to 1, then this case reduces to the preceding cases by the formula $\{x_0, 1\} = \{x_0 + 1\}$. If $x'_1 \neq 1$ and $x''_1 \neq 1$ then the point $x = \{x''_0\}$ separates $x'$ and $x''$.

THE EIGHTH CASE: $\alpha' = 0$, $\alpha'' \geqq 1$, $x'_0 = x''_0$. In this case the point $x = \{x''_0, (x''_1 + 1)\}$ lies between $x' = \{x'_0\}$ and $x'' = \{x'_0, x''_1, x''_2, \cdots, x''_{\alpha''}\}$.

THE NINTH CASE: $\alpha' = \alpha'' = 1$, $x'_0 = x''_0$. In this case the point $x = \{x''_0, x''_1, 2\}$ lies between $x' = \{x'_0, x'_1\}$ and $x'' = \{x''_0, x''_1\}$ for $x' < x''$ and therefore $x'_1 > x''_1$.

The consideration of all preceding particular cases furnishes us the following result: If either $\alpha' \leqq 1$ and $\alpha'' \leqq 1$, or $\alpha = 0$, then there exists a rational point $x$ which separates $x'$ and $x''$. Now let us return to the consideration of the more general cases.

THE FIRST GENERAL CASE: $\alpha \geqq 1$, $x'_0 < x''_0$. In this case the point $x = \{x''_0\}$ lies between $x'$ and $x''$.

THE SECOND GENERAL CASE: $\alpha \geqq 1$, $x'_0 = x''_0$. In this case we suppose that $x'_n \neq x''_n$, but $x'_{n_1} = x''_{n_1}$ for all $n_1 < n \leqq \alpha$. Then we can find a point $\{x_n, x_{n+1}, \cdots, x_{n+\beta}\}$ lying between $\{x'_n, x'_{n+1}, \cdots, x'_{\alpha'}\}$ and $\{x''_n, x''_{n+1}, \cdots, x''_{\alpha''}\}$.

See the first general case. It is obvious that the point $x = \{x_0', x_1', x_2', \cdots, x_{n_1}', \cdots, x_n, x_{n+1}, \cdots, x_{n+\beta}\}$ lies between $x'$ and $x''$.

THE THIRD GENERAL CASE: $\alpha \geqq 1$, $x_k' = x_k''$ for all $k = 0, 1, 2, 3, \cdots, \alpha$. In this case we find a point $\{x_\alpha, x_{\alpha+1}, \cdots, x_{\alpha+\beta}\}$ lying between $\{x_\alpha', x_{\alpha+1}', \cdots, x_{\alpha'}'\}$ and $\{x_\alpha'', x_{\alpha+1}'', \cdots, x_{\alpha''}''\}$. See the eighth particular case. It is obvious that the point $x = \{x_0', x_1', x_2', \cdots, x_{\alpha_1}', \cdots, x_\alpha, x_{\alpha+1}, \cdots, x_{\alpha+\beta}\}$ where $\alpha_1 < \alpha$, lies between $x'$ and $x''$. Thus we have considered all the cases, and our theorem is proved.

## The Set $I_x^{(i)}$

Let

$$x = \{x_0, x_1, x_2, \cdots, x_\alpha, \cdots\} \qquad (0 \leqq \alpha < \Omega_i) \tag{2}$$

be a sequence of the transfinite numbers $x_\alpha$, $1 \leqq x_\alpha < \Omega_i$. This sequence will be called an *irrational* point of the transfinite space $E_x^{(i)}$. Denote by $I_x^{(i)}$ the set of all irrational points. It is obvious that $E_x^{(i)} = R_x^{(i)} + I_x^{(i)}$. The rational point

$$\bar{x}_\alpha = \{x_0, x_1, x_2, \cdots, x_\alpha\} \tag{3}$$

is called the rational approximation $\alpha$ for the point

$$x = \{x_0, x_1, x_2, \cdots, x_\alpha, \cdots\} \qquad (\alpha < \Omega_i).$$

We have the following evident proposition on rational approximations:

(i) if $\alpha$ runs over the increasing sequence of the even numbers, then $\bar{x}_\alpha$ also increases;

(ii) if $\alpha$ runs over the increasing sequence of the odd numbers, then $\bar{x}_\alpha$ decreases. For this reason we state the following

DEFINITION 6. *An irrational point* $x = \{x_0, x_1, x_2, \cdots, x_\gamma, \cdots\}$ $(\gamma < \Omega_i)$ *lies to the left of* $\bar{x}_\delta = \{x_0, x_1, x_2, \cdots, x_\delta\}$ *and to the right of* $\bar{x}_\alpha = \{x_0, x_1, x_2, \cdots, x_\alpha\}$ *where* $\alpha$ *is an even number and* $\delta$ *is an odd number.*

DEFINITION 7. *If* $y$ *is any point of* $E_x^{(i)}$ *and if* $\bar{x}_\alpha > y$ $[\bar{x}_\alpha < y]$ *for all odd [even] numbers* $\alpha$, $\alpha < \Omega_i$, *then we have*

$$x \geqq y \; [x \leqq y].$$

The preceding definitions 6 and 7 characterize the relation of inequality between points of $E_x^{(i)}$. In order to define the equality of points of $E_x^{(i)}$ we make the following definitions.

DEFINITION 8. *If a point* $y$ *satisfies the inequality*

$$\bar{x}_\alpha < y < \bar{x}_\delta$$

*for all odd numbers* $\delta$ *and for all even numbers* $\alpha$, *then we put* $y = x$.

DEFINITION 9. *If we have*

$$x_0' = x_0'', x_1' = x_1'', \cdots, x_\alpha' = x_\alpha'', \cdots \qquad (\alpha < \Omega_i), \tag{4}$$

*then the point* $x' = \{x_0', x_1', x_2', \cdots, x_\alpha', \cdots\}$ $(\alpha < \Omega_i)$ *is equal to the point*

$$x'' = \{x_0'', x_1'', x_2'', \cdots, x_\alpha'', \cdots\} \qquad (\alpha < \Omega_i).$$

THEOREM III. *Let $a$ and $b$ be any points of the space $E_x^{(i)}$, then we have either $a = b$, $a < b$, or $a > b$.*

PROOF. We consider the following cases.

THE FIRST CASE:

$$a = \{a_0, a_1, a_2, \cdots, a_\alpha\} \qquad (\alpha < \Omega_i)$$

$$b = \{b_0, b_1, b_2, \cdots, b_\beta, \cdots\} \qquad (\beta < \Omega_i).$$

Let $\gamma$ be an even number and let

$$\bar{b}_\gamma = \{b_0, b_1, b_2, \cdots, b_\gamma\} \qquad (\gamma < \Omega_i).$$

If there exists an even number $\gamma$ such that $a \leqq \bar{b}_\gamma$, then we have $a < b$, since $a \leqq \bar{b}_\gamma < b$. If for all even numbers $\gamma$, $\bar{b}_\gamma < a$, then we consider an approximation

$$\bar{b}_\delta = \{b_0, b_1, b_2, \cdots, b_\delta\}$$

where $d$ is an odd number. If there exists an odd number $\delta$ such that $\bar{b}_\delta \leqq a$, then $b < a$, since $b < \bar{b}_\delta \leqq a$. If for all odd numbers $\delta$, $a < \bar{b}_\delta$, then, on account of the relation $\bar{b}_\gamma < a$, we obtain $\bar{b}_\gamma < a < \bar{b}_\delta$ where $\gamma$ is an arbitrary even number and $\delta$ is an arbitrary odd number. Then by the definition 6 we have the equality $a = b$ which contradicts the fact that $a$ is a rational and $b$ an irrational point.

THE SECOND CASE:

$$a = \{a_0, a_1, a_2, \cdots, a_\alpha, \cdots\} \qquad (\alpha < \Omega_i)$$

$$b = \{b_0, b_1, b_2, \cdots, b_\beta, \cdots\} \qquad (\beta < \Omega_i).$$

In this case we can introduce the designations

$$\bar{a}_\alpha = \{a_0, a_1, a_2, \cdots, a_\alpha\}$$

$$\bar{b}_\beta = \{b_0, b_1, b_2, \cdots, b_\beta\}.$$

Let $\alpha_1$, $\beta_1$ be any odd numbers and let $\alpha_2$, $\beta_2$ be any even numbers. If there exist numbers $\alpha_1$, $\beta_2$ such that $\bar{a}_{\alpha_1} \leqq \bar{b}_{\beta_2}$, then we have $a < b$, because $a < \bar{a}_{\alpha_1} \leqq \bar{b}_{\beta_2} < b$. If for all the numbers $\alpha_1$, $\beta_2$ the inequality $\bar{a}_{\alpha_1} > \bar{b}_{\beta_2}$ holds, then we shall consider the points $\bar{a}_{\alpha_2}$ and $\bar{b}_{\beta_1}$. If there exist numbers $\alpha_2$ and $\beta_1$ such that $\bar{b}_{\beta_1} \leqq \bar{a}_{\alpha_2}$, then we have $b < a$, because $b < \bar{b}_{\beta_1} \leqq \bar{a}_{\alpha_2} < a$. If for all the numbers $\alpha_2$ and $\beta_1$ the inequality $\bar{b}_{\beta_1} > \bar{a}_{\alpha_2}$ holds, then, on account of the relation $\bar{a}_{\alpha_1} > \bar{b}_{\beta_2}$, we obtain $\bar{b}_{\beta_1} > \bar{a}_{\alpha_2}$, $\bar{a}_{\alpha_1} > \bar{b}_{\beta_2}$. Thence by definition 7 we have $\bar{b}_{\beta_1} > a$, $a > \bar{b}_{\beta_2}$. This means that $\bar{b}_{\beta_2} < a < \bar{b}_{\beta_1}$. From this relation by definition 6 we obtain $a = b$. Thus our theorem is proved.

THEOREM IV. *Let $a$ and $b$ be any points of the space $E_x^{(i)}$ such that $a < b$, then there exists a rational point $c$ of $E_x^{(i)}$ such that $a < c < b$, that is to say, the set $E_x^{(i)}$ is chained.*

PROOF. We shall consider the following cases.

THE FIRST CASE:

$$a = \{a_0, a_1, a_2, \cdots, a_\alpha\} \qquad (\alpha < \Omega_i)$$

$$b = \{b_0, b_1, b_2, \cdots, b_\beta, \cdots\} \qquad (\beta < \Omega_i).$$

Since $a < b$ there exists an even number $\beta_2$ such that $a \leqq \bar{b}_{\beta_2}$. Indeed, if we assume that the inequality $a > \bar{b}_{\beta_2}$ holds for all even numbers $\beta_2$, then by definition 7 we have $a \geqq b$ which is impossible. Thus we have $a \leqq \bar{b}_{\beta_2} < \bar{b}_{\beta_2+2} < b$. Let $c$ be a rational point such that $\bar{b}_{\beta_2} < c < \bar{b}_{\beta_2+2}$, then we have $a < c < b$.

THE SECOND CASE:

$$a = \{a_0, a_1, a_2, \cdots, a_\alpha, \cdots\} \qquad (\alpha < \Omega_i)$$

$$b = \{b_0, b_1, b_2, \cdots, b_\beta\} \qquad (\beta < \Omega_i).$$

Since $a < b$ there exists an odd number $\alpha_1$ such that $a_{\alpha_1} \leqq b$. Indeed, if we assume that the inequality $a_{\alpha_1} > b$ holds for all odd numbers $\alpha_1$, then by definition 7 we have $a \geqq b$ which is impossible. Thus we have $a < a_{\alpha_1} < a_{\alpha_1+2} \leqq b$. Let $c$ be a rational point such that $a_{\alpha_1} < c < a_{\alpha_1+2}$, then we have $a < c < b$.

THE THIRD CASE:

$$a = \{a_0, a_1, a_2, \cdots, a_\alpha, \cdots\} \qquad (\alpha < \Omega_i)$$

$$b = \{b_0, b_1, b_2, \cdots, b_\beta, \cdots\} \qquad (\beta < \Omega_i).$$

Since $a < b$ there exists an even number $\beta_2$ and an odd number $\alpha_1$ such that $\bar{a}_{\alpha_1} \leqq \bar{b}_{\beta_2}$. Indeed, if we assume that the inequality $\bar{a}_{\alpha_1} > \bar{b}_{\beta_2}$ holds for all $\alpha_1$ and $\beta_2$, then by applying the reasoning of theorem III we can show that $a \geqq b$, but this is impossible. Thus $\bar{a}_{\alpha_1} \leqq \bar{b}_{\beta_2}$ from whence it follows $a < \bar{a}_{\alpha_1} \leqq \bar{b}_{\beta_2} < \bar{b}_{\beta_2+2} < b$. Let $c$ be a rational point such that $\bar{b}_{\beta_2} < c < \bar{b}_{\beta_2+2}$, then we have $a < c < b$.

THE FOURTH CASE: the points $a$ and $b$ are rational.

In this case by theorem II there exists a rational point $c$ lying between $a$ and $b$. Thus our theorem is proved, and the set $E_x^{(i)}$ is chained.

THEOREM V. *If the totality $R_x^{(i)}$ of all rational points is separated into two distinct classes $X$, $Y$ such that every point of $X$ lies to the left of each point of $Y$, then there exists an irrational point $i$ such that $x < i < y$ where $x$ is any point of $X$ and $y$ is any point of $Y$.*

PROOF. Let

$$x = \{x_0, x_1, x_2, \cdots, x_\alpha\} \qquad (\alpha < \Omega_i)$$

$$y = \{y_0, y_1, y_2, \cdots, y_\beta\} \qquad (\beta < \Omega_i)$$

and let $M_0$ be the least of the numbers $y_0$. If $M_0$ is either 0 or a number of the second species, then every number $M$, $M \leqq M_0$, belongs to $X$ and the number $(M_0 + 1)$ belongs to $Y$. If $M_0$ is a number of the first species, then $(M_0 + 1)$ belongs to $Y$. We first consider all rational points

$$z_1 = \{M_0, x_1, x_2, \cdots, x_\alpha\} \qquad (\alpha < \Omega_i)$$

of the class $X$. Let $M_1$ be the least of the numbers $x_1$. It is clear that every point $\{M_0, M_1'\}$, where $M_1' \leqq M_1$, belongs to $Y$, and that the point $\{M_0, (M_1 + 1)\}$ belongs to $X$. Next we consider all rational points

$$z_2 = \{M_0, M_1, x_2, x_3, \cdots, x_\alpha\} \qquad (\alpha < \Omega_i)$$

of $Y$. Denote by $M_2$ the least of the numbers $x_2$. It is clear that every point $\{M_0, M_1, M_2'\}$, where $M_2' \leqq M_2$, belongs to $Y$, and that the point $\{M_0, M_1, (M_2 + 1)\}$ belongs to $X$, and so on. This procedure furnishes us the irrational point

$$M = \{M_0, M_1, M_2, \cdots, M_\alpha, \cdots\} \qquad (\alpha < \Omega_i).$$

It can be verified without difficulty that $x < M < y$. Indeed, we have

$$\{x_0, x_1, x_2, \cdots, x_{\alpha_2}\} \leqq \{M_0, M_1, M_2, \cdots, M_{\alpha_2}\}$$

$$\{M_0, M_1, M_2, \cdots, M_{\beta_1}\} \leqq \{y_0, y_1, y_2, \cdots, y_{\beta_1}\}$$

for every even number $\alpha_2$ and for every odd number, since

$$x_0 < M_0 \leqq y_0$$

$$y_1 < M_1 \leqq x_1$$

$$x_2 < M_2 \leqq y_2$$

$$\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots$$

THEOREM VI. *If the set $E_x^{(i)} + \xi$ which we obtain by adjoining to the set $E_x^{(i)}$ a new element $\xi$ is ordered and chained, then $\xi$ is either the greatest or the least element of the set $E_x^{(i)} + \xi$.*

PROOF. In order to prove this theorem we assume that $\xi$ is neither the greatest nor the least element of the ordered and chained set $E_x^{(i)} + \xi$. It is clear that if $l$ is any point of $E_x^{(i)}$, then either $l < \xi$ or $l > \xi$. This means that the totality of all points of $E_x^{(i)}$ can be separated into two distinct classes $X_1$ and $Y_1$. The class $X_1$ consists of all the points $l$ for which the inequality $l < \xi$ holds. The class $Y_1$ consist of all the points $l$ for which the inequality $l > \xi$ holds. By hypothesis, $\xi$ cannot be either the greatest or the least element of $E_x^{(i)}$, consequently, each one of the sets $X_1$ and $Y_1$ contains at least one element. Besides, every point of $X_1$ lies to the left of each point of $Y_1$.

Here we must consider the following cases:

1) $X_1$ has a greatest element and $Y_1$ has no least element;
2) $X_1$ has no greatest element and $Y_1$ has a least element;

3) $X_1$ has a greatest element and $Y_1$ has a least element;
4) $X_1$ has no greatest element and $Y_1$ has no least element.

Case 3) is impossible, because we can find a point $c$ of $E_x^{(i)}$ lying between the greatest element of $X_1$ and the least element of $Y_1$, and hence not belonging to $E_x^{(i)}$, a contradiction.

We next consider case 4). Let $X$ be the set of all rational points of $X_1$ and let $Y$ be the set of all rational points of $Y_1$. It is clear that $R_x^{(i)} = X + Y$ and that every point of $X$ lies to the left of each point of $Y$. We now suppose a) that $X$ has a greatest element $M$ and $Y$ has a least element $m$. Then there exists a point $c$ of $R_x^{(i)}$ lying between $M$ and $m$ which, therefore, does not belong to $R_x^{(i)}$.

Next we suppose b) that $X$ has no greatest element and $Y$ has no least element. In this case there exists an irrational point $i$ such that $x < i < y$.

Let $i \in X_1$. Then, if $x_1 \in X_1$, there exists a point $\bar{x}_1$ of $X_1$ such that $x_1 < \bar{x}_1$. Let $c$ be a rational point lying between $x_1$ and $\bar{x}_1$. Then $c$ belongs to $X_1$ and consequently to $X$, therefore $c < i$ and $x_1 < i$. This means that $i$ is the greatest element of $X_1$, but this is impossible in case 4). Let $i \in Y_1$. In this case the same reasoning as above proves that $i$ is the least element of $Y$, but this is impossible in case 4). Thereupon, we suppose c) that $X$ has a greatest element, $M$, and $Y$ has no least element. It is clear that $M$ is the greatest element of $X_1$. Indeed, if $M$ is not the greatest element of $X_1$, then there exists a point $H$ of $X_1$ such that $M < H$ and therefore there exists a rational point $c$ such that $M < c < H$, consequently, $c$ belongs to $X$, which is impossible.

Thus, $X_1$ has a greatest element, but in case 4) $X_1$ has no greatest element. Finally, we suppose d) that $Y$ has a least element, $m$. By the same reasoning we can prove that, if $m$ is the least element of $Y$, $m$ is also the least element of $Y_1$, this is impossible in case 4). Thus cases 3) and 4) are impossible.

In case 1) there exists a greatest element, $M$, of $X_1$. Since $M < \xi$ and there is no point of the set $(E_x^{(i)} + \xi)$ lying between $M$ and $\xi$, this set is then not chained. This contradicts our supposition that the set $(E_x^{(i)} + \xi)$ is chained. In case 2) the same reasoning leads to a contradiction. Thus our theorem is proved.

THEOREM VI. *The power of the set $E_x^{(i)}$ is equal to $2^{\aleph_i}$.*

This theorem is obvious.

It is important to notice here that, if we denote by $\{a_0, a_1, a_2, \cdots, a_{\Omega_0}\}$, where $a_{\Omega_0} = a_0$, the continued fraction

$$a_0 + \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \ddots,$$

then the set of all these fractions is contained in $E_x^{(i)}$, $E_x^{(0)} \subset E_x^{(i)}$.

TCHEBOKSARY, U.S.S.R.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# CONTINUED FRACTION EXPANSIONS FOR ARBITRARY POWER SERIES

BY W. T. SCOTT AND H. S. WALL

(Received March 23, 1939)

An arbitrary power series $P(x) = \sum_{n=0}^{\infty} c_n x^n$ has a unique corresponding continued fraction of the form

$$K(x) = c_0 + \frac{a_1 x^{\alpha_1}}{1} + \frac{a_2 x^{\alpha_2}}{1} + \frac{a_3 x^{\alpha_3}}{1} + \cdots$$

which terminates if and only if $P(x)$ represents a rational function of $x$. A fundamental problem is to obtain formulas for the coefficients $a_n$ and the exponents $\alpha_n$ in terms of the given power series. This the authors have done in the present paper for a large class of *absolutely regular* power series [Definition 2.2]. These include among others all power series of the form $1 + \sum_1^{\infty} c_n x^{\lambda_{n-1}}$, where $\lambda_n \geqq 2\lambda_{n-1}$, $c_n \neq 0$, $n = 1, 2, 3, \cdots$, and the reciprocals of these series when $\lambda_n > 2\lambda_{n-1}$, $n = 1, 2, 3, \cdots$. The class of absolutely regular power series contains the class of *semi-normal* power series [Perron (1), p. 304], and the formulas found are natural extensions of the well-known formulas for the semi-normal case.

Another question which arises concerns the relationship of the convergents of $K(x)$ to the Padé table of rational approximants for $P(x)$ [Perron (1), Chapter 10]. It turns out that this question is closely related to that of determining simple formulas for the partial numerators of the continued fraction.

The authors conclude with a discussion of a class of functions for which the existence of a natural boundary can be argued by means of the corresponding continued fractions.

## 1. Corresponding and associated continued fractions

A continued fraction of the form

$$1 + \frac{a_1 x^{\alpha_1}}{1} + \frac{a_2 x^{\alpha_2}}{1} + \frac{a_3 x^{\alpha_3}}{1} + \cdots \tag{1.1}$$

in which $a_1, a_2, a_3, \cdots$ are real or complex numbers different from 0, and $\alpha_1, \alpha_2, \alpha_3, \cdots$ are positive integers, will be called a *corresponding type* continued fraction. Let $A_n(x)/B_n(x)$, $n = 0, 1, 2, 3, \cdots$, be the sequence of convergents of (1.1), where $A_0(x) = 1$, $B_0(x) = 1$, $A_1(x) = 1 + a_1 x^{\alpha_1}$, $B_1(x) = 1$, and

$$\begin{aligned} A_n(x) &= A_{n-1}(x) + a_n x^{\alpha_n} A_{n-2}(x) \\ B_n(x) &= B_{n-1}(x) + a_n x^{\alpha_n} B_{n-2}(x) \end{aligned} \tag{1.2}$$

$n = 2, 3, 4, \cdots$. Then from the well-known identity [Perron (1), p. 16]:

$$\frac{A_n(x)}{B_n(x)} - \frac{A_{n+1}(x)}{B_{n+1}(x)} = (-1)^{n+1} \frac{a_1 a_2 \cdots a_{n+1}}{B_n(x) B_{n+1}(x)} x^{\alpha_1+\alpha_2+\cdots+\alpha_{n+1}} \tag{1.3}$$

and from the fact that $B_n(0) = 1$, it follows that there exists uniquely a power series $P(x) = 1 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$ such that

$$B_n(x)P(x) - A_n(x) = (x^{\alpha_1+\alpha_2+\cdots+\alpha_{n+1}}) \tag{1.4}$$

where $(x^s)$ denotes a formal power series in which the $s^{\text{th}}$ but no lower power of $x$ appears. We call $P(x)$ the power series which *corresponds* to (1.1). It is impossible for $P(x)$ to represent a rational function of $x$. For let $P(x) \equiv N(x)/D(x)$ where $N(x)$ and $D(x)$ are polynomials. Then for every $n$:

$$N(x)B_n(x) - D(x)A_n(x) = (x^{\alpha_1+\alpha_2+\cdots+\alpha_n+1}).$$

Now the degree of the left member is at most the greater of the degrees of $N(x)B_n(x)$ and $D(x)A_n(x)$. But the differences $(\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{n+1} - s_n)$ and $(\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{n+1} - t_n)$, where $s_n$ and $t_n$ are the degrees of $A_n(x)$ and $B_n(x)$, approach $\infty$ with $n$. In fact, it is easy to show by induction that the first difference is $\geqq [n/2]$, where $[x]$ denotes the largest integer contained in $x$; and a similar inequality holds for the second difference. This contradiction shows that the power series corresponding to (1.1) cannot represent a rational function of $x$.

Suppose conversely that $P(x) = 1 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$ is a power series which does not represent a rational function of $x$. If $c_i = 0$, $i = 1, 2, 3, \cdots$, $\alpha_1 - 1, c_{\alpha_1} \neq 0$, we take $a_1 = c_{\alpha_1}$, $A_0(x) = B_0(x) = 1$, $A_1(x) = 1 + a_1 x^{\alpha_1}$, $B_1(x) = 1$, and determine $a_2 \neq 0$, $\alpha_2 > 0$, by the condition

$$\frac{A_1(x)}{B_1(x)} - P(x) = a_1 a_2 x^{\alpha_1+\alpha_2} + (x^s), \qquad s > \alpha_1 + \alpha_2.$$

This is possible since the difference on the left cannot vanish identically. We then put $A_2(x) = A_1(x) + a_2 x^{\alpha_2} A_0(x)$, $B_2(x) = B_1(x) + a_2 x^{\alpha_2} B_0(x)$, and determine $a_3 \neq 0$, $\alpha_3 > 0$, by the condition

$$\frac{A_2(x)}{B_2(x)} - P(x) = -a_1 a_2 a_3 x^{\alpha_1+\alpha_2+\alpha_3} + (x^s), \quad s > \alpha_1 + \alpha_2 + \alpha_3.$$

Having by this process determined $a_i \neq 0$, $\alpha_i > 0$, $A_i(x)$, $B_i(x)$, we determine $a_{i+1} \neq 0$, $\alpha_{i+1} > 0$, and hence $A_{i+1}(x)$, $B_{i+1}(x)$, from the relation

$$\frac{A_i(x)}{B_i(x)} - P(x) = (-1)^{i+1} a_1 a_2 \cdots a_{i+1} x^{\alpha_1+\alpha_2+\cdots+\alpha_{i+1}} + (x^s)$$

where $s > \alpha_1 + \alpha_2 + \cdots + \alpha_{i+1}$. We thus determine inductively an infinite continued fraction (1.1) such that the $n^{\text{th}}$ convergent satisfies (1.4) for $n = 0, 1, 2, \cdots$.

Suppose now that there exist two different continued fractions (1.1) and

$$1 + \frac{a_1 x^{\alpha_1}}{1} + \frac{a_2 x^{\alpha_2}}{1} + \cdots + \frac{a_{n-1} x^{\alpha_{n-1}}}{1} + \frac{b_n x^{\beta_n}}{1} + \cdots \tag{1.5}$$

differing first in their $n^{\text{th}}$ partial quotients, both having the corresponding power series $P(x)$. Since their $(n-1)^{\text{th}}$ convergents are the same, the expansions into power series must agree with $P(x)$ for the same number of consecutive terms beginning with the first. This implies that $\alpha_n = \beta_n$. Let $A'_n(x)/B'_n(x)$ be the $n^{\text{th}}$ convergent of (1.5). Then

$$B'_n(x)P(x) - A'_n(x) = (x^{\alpha_1+\alpha_2+\cdots+\alpha_n+\beta_{n+1}}),$$
$$B_n(x)P(x) - A_n(x) = (x^{\alpha_1+\alpha_2+\cdots+\alpha_n+\alpha_{n+1}}).$$

This implies that

$$\frac{A'_n(x)}{B'_n(x)} - \frac{A_n(x)}{B_n(x)} = (x^{\alpha_1+\alpha_2+\cdots+\alpha_n+c})$$

where $c > 0$. But this difference is a formal power series in which the degree of the lowest degree term which appears is $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ unless $a_n = b_n$. We have proved the following theorem.

THEOREM 1.1. *There is a one to one correspondence between non-terminating continued fractions of the form* (1.1) *and power series* $P(x) = 1 + \sum_{i=1}^{\infty} c_i x^i$ *which do not represent rational functions of* $x$. *This correspondence is completely determined by the condition* (1.4).

In proving this theorem we have set forth a fairly practical step by step procedure for actually expanding a power series into its corresponding continued fraction.

From the proof given by Leighton and Scott [Leighton and Scott (2), p. 598] one obtains the following compound determinant expression for the coefficient $a_n$ in (1.1) in terms of the coefficients of $P(x)$, namely: put

$$c_n^{(i)} = \frac{(-1)^n}{[c_{\alpha_i}^{(i-1)}]^n} \begin{vmatrix} c_{1+\alpha_i}^{(i-1)} & c_{\alpha_i}^{(i-1)} & 0 & \cdots & 0 \\ c_{2+\alpha_i}^{(i-1)} & c_{1+\alpha_i}^{(i-1)} & c_{\alpha_i}^{(i-1)} & 0 \cdots & 0 \\ & \cdot & \cdot & \cdot & 0 \\ & & & & c_{\alpha_i}^{(i-1)} \\ c_{n+\alpha_i}^{(i-1)} & c_{n+\alpha_i+1}^{(i-1)} & & \cdots & c_{1+\alpha_i}^{(i-1)} \end{vmatrix}. \tag{1.6}$$

$i = 1, 2, 3, \cdots, n = 1, 2, 3, \cdots$, where $c_{\alpha_i}^{(i-1)} \neq 0$, $i = 1, 2, 3, \cdots$, $c_n^{(0)} = c_n$, and $c_n^{(i-1)} = 0$, $n = 1, 2, 3, \cdots, \alpha_i - 1$. Then

$$a_i = c_{\alpha_i}^{(i-1)}, \qquad i = 1, 2, 3, \tag{1.7}$$

In fact, the power series $P_i(x) = 1 + \sum_{n=1}^{\infty} c_n^{(i)} x^n$, $i = 0, 1, 2, \cdots$, $P_0(x) = P(x)$, satisfy the formal power series identities

$$P_i(x) = 1 + \frac{a_{i+1}x^{\alpha_{i+1}}}{P_{i+1}(x)} \tag{1.8}$$

and these identities generate the corresponding continued fraction. As a consequence we have the following theorem.

THEOREM 1.2. *Any two of the following relations implies the third*:

$$\text{(a)} \qquad P(x) \sim 1 + \frac{a_1x^{\alpha_1}}{1} + \frac{a_2x^{\alpha_2}}{1} + \frac{a_3x^{\alpha_3}}{1} + \cdots,$$

$$\text{(b)} \qquad P_1(x) \sim 1 + \frac{a_2x^{\alpha_2}}{1} + \frac{a_3x^{\alpha_3}}{1} + \frac{a_4x^{\alpha_4}}{1} + \cdots,$$

$$\text{(c)} \qquad P(x) = 1 + \frac{a_1x^{\alpha_1}}{P_1(x)}$$

From (1.3) and the fact that $A_n(0) = B_n(0) = 1$ we have this theorem:

THEOREM 1.3. *The $n^{\text{th}}$ convergent of a corresponding type continued fraction is an irreducible rational fraction.*

In order to obtain an extension of the notion of "associated type" continued fractions [Perron (1), p. 322], we contract (1.1) to form the continued fraction having for its sequence of convergents the sequence of even convergents of (1.1) namely [Perron (1), p. 201]:

$$1 + \frac{a_1x^{\alpha_1}}{(1 + a_2x^{\alpha_2})} - \frac{a_2a_3x^{\alpha_2+\alpha_3}}{(1 + a_3x^{\alpha_3} + a_4x^{\alpha_4})} - \frac{a_4a_5x^{\alpha_4+\alpha_5}}{(1 + a_5x^{\alpha_5} + a_6x^{\alpha_6})} - \cdots. \tag{1.9}$$

By means of a sequence of power series identities this continued fraction can always be constructed in a unique fashion directly from the power series without first constructing (1.1). To do this, let $c_{\alpha_1}x^{\alpha_1}$ be the non-vanishing term of lowest degree in $P(x) - 1$, put $a_1 = c_{\alpha_1}$ and determine the power series $P^{(1)}(x) = 1 + c_1^{(1)}x + c_2^{(1)}x^2 + \cdots$ to satisfy the following power series identity:

$$P(x) = 1 + \frac{a_1x^{\alpha_1}}{P^{(1)}(x)}.$$

Let $c_1^{(1)} = c_2^{(1)} = \cdots = c_{\alpha_2-1}^{(1)} = c_{\alpha_2+1}^{(1)} = \cdots = c_{\alpha_2+\alpha_3-1}^{(1)} = 0$, $c_{\alpha_2}^{(1)} \neq 0$, $c_{\alpha_2+\alpha_3}^{(1)} \neq 0$, and take $a_2 = c_{\alpha_2}^{(1)}$, $-a_2a_3 = c_{\alpha_2+\alpha_3}^{(1)}$. This determines $a_1$, $a_2$, $a_3$, $\alpha_1$, $\alpha_2$, $\alpha_3$ in (1.9). Next determine $P^{(2)}(x) = 1 + c_1^{(2)}x + c_2^{(2)}x^2 + \cdots$ such that

$$P^{(1)}(x) = 1 + a_2x^{\alpha_2} - \frac{a_2a_3x^{\alpha_2+\alpha_3}}{P^{(2)}(x)}.$$

Write $P^{(2)}(x) = 1 + a_3x^{\alpha_3} + Q^{(2)}(x)$ and determine $a_4x^{\alpha_4}$ as the non-vanishing term of lowest degree in $Q^{(2)}(x)$, and $-a_4a_5x^{\alpha_4+\alpha_5}$ as the next lowest degree non-

vanishing term. This determines $a_4$, $a_5$, $\alpha_4$, $\alpha_5$. We then determine $P^{(3)}(x) = 1 + c_1^{(3)}x + c_2^{(3)}x^2 + \cdots$ such that

$$P^{(2)}(x) = 1 + a_3x^{\alpha_3} + a_4x^{\alpha_4} - \frac{a_4a_5x^{\alpha_4+\alpha_5}}{P^{(3)}(x)}$$

and continue the process. One may readily verify that the continued fraction thus constructed has for its sequence of convergents the sequence of even convergents of the continued fraction corresponding to $P(x)$.

Conversely, if we are given a continued fraction of the form (1.9) with $n^{\text{th}}$ convergent $L_n(x)/M_n(x)$, then there is uniquely determined a power series $P(x)$ such that $P(x)M_n(x) - L_n(x) = (x^{\alpha_1+\alpha_2+\alpha_3+\cdots+\alpha_{2n+1}})$, $n = 0, 1, 2, 3, \cdots$. We shall say that (1.9) and $P(x)$ are *associated*.

In the special case where $a_{2n+1} = -a_{2n+2}$, $\alpha_{2n+1} = \alpha_{2n+2}$ $n = 1, 2, 3, \cdots$ in (1.9), we have, on changing the notation, the following theorem.

THEOREM 1.4. *The correspondence*

$$P(x) \sim 1 + \frac{a_1x^{\alpha_1}}{1} + \frac{a_2x^{\alpha_2}}{1} - \frac{a_2x^{\alpha_2}}{1} + \frac{a_3x^{\alpha_3}}{1} - \frac{a_3x^{\alpha_3}}{1} + \cdots$$

*implies the correspondence*

$$P(x) - c_{\alpha_1}x^{\alpha_1} \sim 1 - \frac{a_1a_2x^{\alpha_1+\alpha_2}}{1} + \frac{a_2a_3x^{\alpha_2+\alpha_3}}{1} + \frac{a_3a_4x^{\alpha_3+\alpha_4}}{1} + \cdots.$$

## 2. Padé approximants and continued fractions of corresponding type

The *Padé table* [Perron (1), Chapter 10] for a power series $P(x) = 1 + \sum_{i=1}^{\infty} c_ix^i$ is a table of double entry of the form

| | | | |
|---|---|---|---|
| (0,0) | (0,1) | (0,2) | ⋯ |
| (1,0) | (1,1) | (1,2) | ⋯ |
| (2,0) | (2,1) | (2,2) | ⋯ |
| . | . | . | . |

in which the *field* $(m, n)$ in the $(m + 1)^{\text{th}}$ row and $(n + 1)^{\text{th}}$ column is occupied by the unique rational fraction $N_{m,n}(x)/D_{m,n}(x)$ in which $D_{m,n}(x) \not\equiv 0$, the degree of $N_{m,n}(x)$ is $\leqq n$, the degree of $D_{m,n}(x)$ is $\leqq m$, and

$$P(x)D_{m,n}(x) - N_{m,n}(x) = (x^s), \qquad s \geqq m + n + 1. \tag{2.1}$$

If $N_{m,n}(x)/D_{m,n}(x)$ is irreducible, $N_{m,n}(0) = D_{m,n}(0) = 1$, and the degrees of numerator and denominator are exactly $n$ and $m$, respectively, then $s = m + n + r + 1$ in (2.1) where $r \geqq 0$, and this Padé approximant occupies the $(r + 1)^2$ fields

$$(m + i, n + j), \qquad i, j = 0, 1, 2, \cdots r$$

and no others, in the Padé table.

Let us now turn to the corresponding continued fraction for $P(x)$ and recall that the $n^{\text{th}}$ convergent $A_n(x)/B_n(x)$ is an irreducible rational fraction such that $A_n(0) = B_n(0) = 1$. Let $s_n$, $t_n$ be the degrees of $A_n(x)$, $B_n(x)$, respectively. Then from (1.4) we see that $A_n(x)/B_n(x)$ is a Padé approximant for $P(x)$ if and only if $\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_{n+1} > s_n + t_n$. It is easy to construct examples to show that this inequality need not hold.

Suppose that

$$s_n + t_n + r_n + 1 = \alpha_1 + \alpha_2 + \cdots + \alpha_{n+1}, \qquad r_n \geqq 0.$$

Then $A_n(x)/B_n(x)$ is a Padé approximant occupying the $(r_n + 1)^2$ fields $(t_n + i, s_n + j)$, $i, j = 0, 1, 2, \cdots, r_n$, in the Padé table, and only these. We shall call this square of equal approximants the *square of index n for* $P(x)$. The set of all the indexed squares of $P(x)$ make up the *corresponding polygon* of $P(x)$.

The convergents $A_0(x)/B_0(x)$, $A_1(x)/B_1(x)$ are always Padé approximants, so that $P(x)$ has indexed squares of index 0 and 1. If $A_2(x)/B_2(x)$ is a Padé approximant, then the first three indexed squares are situated in this way: the square of index $2n + 1$ lies to the right of the square of index $2n$ and they have a portion of a side in common; and the square of index $2n + 2$ lies below the square of index $2n + 1$, and they have a portion of a side in common.

DEFINITION 2.1. The $n^{\text{th}}$ convergent of a corresponding type continued fraction is *regular* if it is a Padé approximant for the corresponding power series. The continued fraction is regular if all its convergents are regular. A power series is regular if its corresponding continued fraction is regular.

We shall refer to the diagonal of an indexed square which is perpendicular to the principal diagonal of the Padé table as the *characteristic diagonal* of the square. If the square is of index $n$ the characteristic diagonal passes through the fields $(i, j)$ of the Padé table for which $i + j = s_n + t_n + r_n$. Consider two indexed squares $Q_n$ and $Q_{n+1}$ with consecutive indices $n$, $n + 1$. Inasmuch as $s_{n+1} + t_{n+1} + r_{n+1} > s_n + t_n + r_n$, we see that the characteristic diagonal of $Q_{n+1}$ is more advanced in the table than that of $Q_n$. Since $A_n(x)/B_n(x) \not\equiv A_{n+1}(x)/B_{n+1}(x)$, the squares $Q_n$ and $Q_{n+1}$ cannot overlap. Hence it follows that one (or both) of the following inequalities must hold:

$$s_{n+1} > s_n + r_n, \tag{2.2}$$

$$t_{n+1} > t_n + r_n. \tag{2.3}$$

If (2.2) holds it follows from (1.2) that

$$s_{n+1} = s_{n-1} + \alpha_{n+1}, \tag{2.4}$$

and if (2.3) holds,

$$t_{n+1} = t_{n-1} + \alpha_{n+1}. \tag{2.5}$$

Suppose now that the square of index $n - 1$ exists. Then $s_{n-1} + t_{n-1} + r_{n-1} + 1 = \alpha_1 + \alpha_2 + \cdots + \alpha_n$, $r_{n-1} \geqq 0$. Consequently if (2.2) holds we have

$$s_{n+1} - (s_n + r_n) = t_n - (t_{n-1} + r_{n-1}) > 0, \tag{2.6}$$

and if (2.3) then

$$t_{n+1} - (t_n + r_n) = s_n - (s_{n-1} + r_{n-1}) > 0. \tag{2.7}$$

We shall state this result as

THEOREM 2.1. *If $Q_{n-1}$, $Q_n$, $Q_{n+1}$ are three indexed squares of $P(x)$ of consecutive index and*

$$Q_k = (t_k + i, s_k + j) \qquad (i, j = 0, 1, 2, \cdots, r_k),$$

*then at least one of the inequalities* (2.2), (2.3) *must hold; and* (2.2) *implies* (2.4), (2.6), *while* (2.3) *implies* (2.5), (2.7).

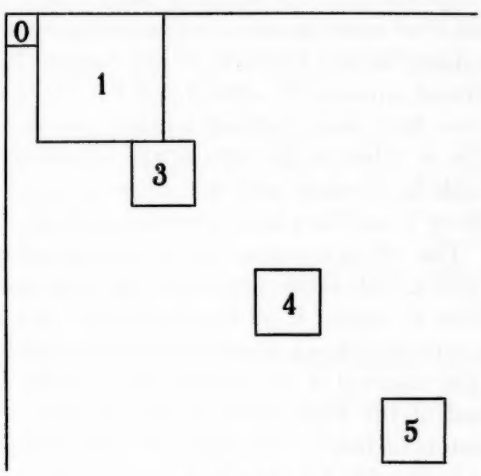



FIG. 1

An example in which both (2.2) and (2.3) hold is afforded by the following continued fraction:

$$1 + x/1 + x^4/1 + x/1 + x^4/1 + x^8/1 + x^8/1 + \cdots.$$

The indexed square of index 2 does not exist. Those of index 0, 1, 3, 4, 5 are shown in Fig. 1.

For regular continued fractions we have the following theorem.

THEOREM 2.2. *If the continued fraction* (1.1) *is regular, then the corresponding polygon of indexed squares is connected, in the sense that any two squares with consecutive indices contain fields with a side in common. Moreover, a square of even index lies below the square of next preceding index, and a square of odd index lies to the right of the square of next preceding index.*

PROOF. Let $Q_0$, $Q_1$, $Q_2$, $\cdots$ be the indexed squares of indices 0, 1, 2, $\cdots$, respectively.

We have seen previously that $Q_0$, $Q_1$, $Q_2$ satisfy the conditions of the theorem.

Suppose that $Q_0, Q_1, \cdots, Q_{2n}$ satisfy the conditions, and consider $Q_{2n+1}$: $(t_{2n+1} + i, s_{2n+1} + j), i, j = 0, 1, 2, \cdots, r_{2n+1}$. Suppose if possible that

$$t_{2n+1} > t_{2n} + r_{2n}. \tag{2.8}$$

Then by Theorem 2.1, $t_{2n+1} - (t_{2n} + r_{2n}) = s_{2n} - (s_{2n-1} + r_{2n-1})$. But by hypothesis the quantity on the right of this last equation is negative, so that (2.8) is impossible. Hence by Theorem 2.1 we must have $s_{2n+1} > s_{2n} + r_{2n}$, and hence $s_{2n+1} - (s_{2n} + r_{2n}) = t_{2n} - (t_{2n-1} + r_{2n-1}) = 1$. Therefore the condi-

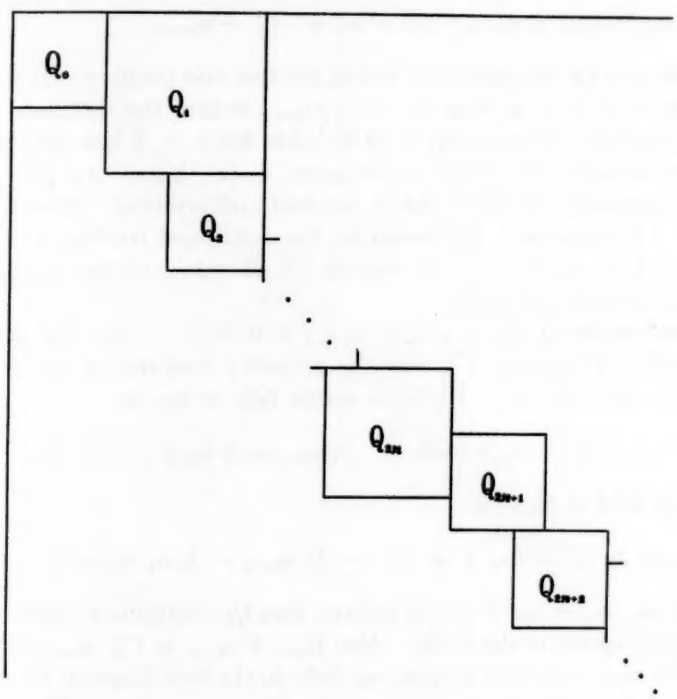



FIG. 2

tions of the theorem are satisfied for $Q_0, Q_1, \cdots, Q_{2n+1}$. In the same way one may show that the conditions are satisfied for $Q_0, Q_1, \cdots, Q_{2n+2}$. This completes the induction.

DEFINITION 2.2. The continued fraction (1.1) and the corresponding power series will be called *absolutely regular* if $\alpha_1 = 1$ and every corresponding type continued fraction with the same sequence of exponents $\alpha_n$ as in (1.1) is regular.

In particular, the semi-normal continued fraction in which $\alpha_n = 1$, $n = 1, 2, 3, \cdots$, is absolutely regular.

THEOREM 2.3. *A necessary and sufficient condition in order that the correspond-*

*ing type continued fraction* (1.1) *be absolutely regular is that the exponents* $\alpha_n$ *satisfy the inequalities*

$$\begin{aligned} &\alpha_2 + \alpha_4 + \cdots + \alpha_{2n} \geqq \alpha_1 + \alpha_3 + \cdots + \alpha_{2n-1}, \\ &\alpha_3 + \alpha_5 + \cdots + \alpha_{2n+1} \geqq \alpha_2 + \alpha_4 + \cdots + \alpha_{2n}, \end{aligned} \qquad n = 1, 2, 3, \cdots. \tag{2.9}$$

PROOF. When these inequalities hold we readily find by mathematical induction that

$$\begin{aligned} s_{2n} \leqq\ & t_{2n} = \alpha_2 + \alpha_4 + \cdots + \alpha_{2n}, \\ \alpha_1 + t_{2n-1} \leqq\ & s_{2n-1} = \alpha_1 + \alpha_3 + \cdots + \alpha_{2n-1}, \end{aligned} \qquad n = 1, 2, 3, \cdots, \tag{2.10}$$

where equality holds throughout if the $a_n$ are real and positive. It then follows by (2.9) that $s_n + t_n < \alpha_1 + \alpha_2 + \cdots + \alpha_{n+1}$, so that the continued fraction is absolutely regular. Conversely, if (2.9) holds for $n < k$ but fails to hold for $n = k$, then at least one of the convergents $A_{2k}(x)/B_{2k}(x)$, $A_{2k-1}(x)/B_{2k-1}(x)$ is not a Padé approximant when the $a_n$ are real and positive. Here the requirement $\alpha_1 = 1$ is essential. For example, the continued fraction $1 + a_1x^3/1 + a_2x/1 + a_3x/1 + a_4x/1 + \cdots$ is regular for all values of the $a_n \neq 0$, but the inequalities (2.9) do not hold.

An indexed square $Q_n:(t_n + i, s_n + j)$, $i, j = 0, 1, 2, \cdots, r_n$, for an absolutely regular continued fraction (1.1) has the property that one of the corner fields depends only upon the $\alpha_n$. In fact a corner field of $Q_{2n}$ is:

$$(t_{2n}, s_{2n} + r_{2n}) = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2n}, \alpha_1 + \alpha_3 + \cdots + \alpha_{2n+1} - 1),$$

and a corner field of $Q_{2n+1}$ is:

$$(t_{2n+1} + r_{2n+1}, s_{2n+1}) = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2n+2} - 1, \alpha_1 + \alpha_3 + \cdots + \alpha_{2n+1}).$$

Since $t_{2n} \geqq s_{2n}$, $s_{2n} + r_{2n} \geqq t_{2n}$, it follows that $Q_{2n}$ contains at least one field in the principal diagonal of the table. Also $t_{2n+1} + r_{2n+1} + 1 \geqq s_{2n+1}$, $t_{2n+1} < s_{2n+1}$, and therefore $Q_{2n+1}$ contains at least one field in the first diagonal file to the right of and parallel to the principal diagonal. Consequently, by Theorem 2.2 we have:

THEOREM 2.4. *The corresponding polygon for an absolutely regular power series contains the fields*

$$(i, i), \qquad (i, i + 1) \qquad (i = 0, 1, 2, \cdots),$$

*of the Padé table.*

This theorem does not hold for the continued fraction $1 + a_1x^3/1 + a_2x/1 + a_3x/1 + a_4x/1 + \cdots$ although it is regular for all $a_n \neq 0$.

If $P_1(x) = 1 + c_2x + c_3x^2 + c_4x^3 + \cdots$ is absolutely regular and $K_1(x)$ is its corresponding continued fraction, then one may readily verify that the convergents of the continued fraction $1 + c_1x + x[K_1(x) - 1]$ are Padé approxi-

mants of $P(x)$ filling the files $(i, i+1)$, $(i, i+2)$, $i = 0, 1, 2, \cdots$, in the Padé table. More generally, we have this theorem:

THEOREM 2.5. *Let the power series* $P_j(x) = 1 + \sum_{i=1}^{\infty} c_{i+j}x^i$, $j = 0, 1, 2, \cdots$, *be absolutely regular and have corresponding continued fractions* $K_j(x)$, $j = 0, 1, 2, \cdots$. *Then the convergents of the continued fractions*

$$1 + c_1x + c_2x^2 + \cdots + c_nx^n + x^n[K_n(x) - 1], \quad n = 0, 1, 2, \cdots, (c_0 = 1),$$

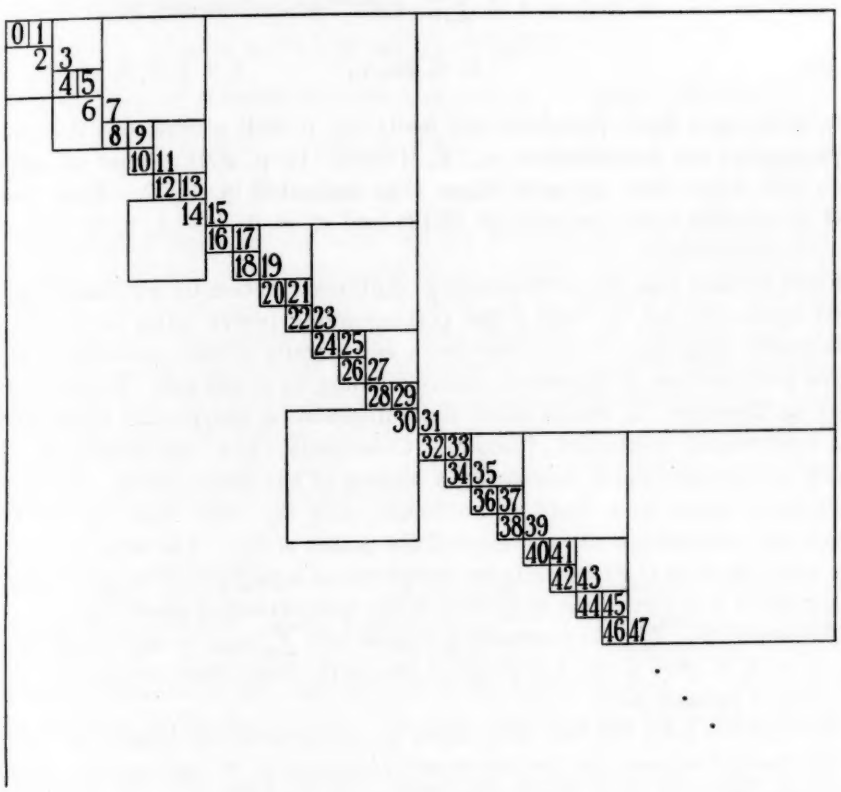



FIG. 3

*are all Padé approximants for* $P_0(x)$, *and their convergents fill all the fields* $(i, j)$, $j \geqq i$, *in the Padé table.*

For semi-normal power series, the sequence of convergents is precisely the sequence of approximants in the fields

$$(0, 0), (0, 1), (1, 1), (1, 2), (2, 2), \cdots, \tag{2.11}$$

and these are all distinct. This does not mean that the indexed squares are all of side unity, as will be seen by examining the Padé table for the semi-normal power series $1 + \sum_{n=0}^{\infty} x^{2^n}$, (Fig. 3).

It is interesting to note that one can tell by inspection of the Padé table the maximum "gap" that a semi-normal power series can have. Let the power series be written in the form $1 + \sum c_i x^{\lambda_i - 1}$ where $1 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots$. Then the section $[1 + c_1 x^{\lambda_0} + c_2 x^{\lambda_1} + \cdots + c_n x^{\lambda_n - 1}]/1$ is a Padé approximant filling the fields $(i, \lambda_{n-1} + j)$, $i, j = 0, 1, 2, \cdots, \lambda_n - \lambda_{n-1}$. Hence since the approximants in the fields (2.11) must all be distinct for a semi-normal power series we see that the following is true:

THEOREM 2.6. *If* $P(x) = 1 + \sum_{i=1}^{\infty} c_i x^{\lambda_i - 1}$ *is semi-normal, then*

$$\lambda_i \leqq 2\lambda_{i-1}, \qquad i = 1, 2, 3, \cdots, \lambda_0 = 1. \tag{2.12}$$

Leighton and Scott [Leighton and Scott (2), p. 604] proved this theorem by showing that the determinants $\varphi_n$, $\psi_n$ [Perron (1), p. 304] are not all different from zero when there are gaps larger than indicated in (2.12). They showed that if equality holds throughout (2.12) and $c_i \neq 0$, $i = 1, 2, 3, \cdots$, then $P(x)$ is semi-normal.

From the fact that the corresponding continued fraction for a rational function must terminate, we see that if the corresponding power series for (1.1) has a sufficiently large gap, there must be a convergent of the continued fraction which is the section of the power series extending up to this gap. For the method used in Theorem 1.1 would show that otherwise a polynomial could have a non-terminating continued fraction. Conversely, any convergent of (1.1) which is a polynomial is necessarily a section of the power series. For such a convergent must be a Padé approximant, and the only Padé approximants which are polynomials are sections of the power series. Likewise we see that any convergent of (1.1) which is the reciprocal of a polynomial is necessarily the reciprocal of a section of the reciprocal of the corresponding power series.

THEOREM 2.7. *If* (1.1) *is absolutely regular and* $\sum(\alpha_{2n+1} - \alpha_{2n}) = \sum(\alpha_{2n+2} - \alpha_{2n+1}) = +\infty$, *then if any two diagonal files of the Padé table converge at* $x = x_0$, *they have a common limit.*

This follows from the fact that under the hypothesis the lengths of the sides of the indexed squares for the power series increase to $\infty$, and any two diagonal files have infinitely many Padé approximants in common.

## 3. Power series in which each exponent is at least twice the preceding

In this section we shall determine completely the corresponding continued fractions for all power series of the form

$$T(x) = c_0 + \sum_{i=1}^{\infty} c_i x^{\lambda_i}, \qquad c_i \neq 0, i = 0, 1, 2, \cdots, \tag{3.1}$$

in which

$$\lambda_{i+1} \geqq 2\lambda_i, \qquad i = 1, 2, 3, \cdots. \tag{3.2}$$

This was done by Leighton and Scott [Leighton and Scott (1), p. 604] for the case in which equality holds throughout in (3.2). Their result may be stated as follows:

THEOREM 3.1. *The power series*

$$c_0 + \sum_{i=1}^{\infty} c_i x^{\lambda_1 2^{i-1}} \qquad (c_i \neq 0,\ i = 0, 1, 2, \cdots, \lambda_1 \geqq 1) \tag{3.3}$$

*has the corresponding continued fraction*

$$c_0 + a_1 x^{\lambda_1}/1 + a_2 x^{\lambda_1}/1 + a_3 x^{\lambda_1}/1 + \cdots \tag{3.4}$$

*where* $a_1, a_2, a_3, \cdots$ *are given in terms of* $c_1, c_2, c_3, \cdots$ *by the following formulas:*

$$\begin{aligned} &a_1 = c_1, \qquad a_{2n} = -a_{2n+1}; \\ &a_{2i+1} = -c_1 c_{i+1}/c_i^2, \quad a_{n\cdot 2^{i+1}+2^{i+1}} = (-1)^n a_{2^{i+1}}, \end{aligned} \qquad i, n = 1, 2, 3, \cdots. \tag{3.5}$$

They obtained this theorem with the aid of a well-known theorem relating to "semi-normal" continued fractions [Perron, p. 304]. We shall begin by showing how the corresponding continued fraction for the general series (3.1) can be derived from the continued fraction (3.4) obtained by Leighton and Scott. We shall find subsequently that it is desirable to give another derivation of the continued fraction for (3.1) which will be independent of Theorem 3.1.

The parameters $c_i$ in (3.3) are different from 0 but otherwise entirely arbitrary. We replace $c_i$ by $c_i x^{\mu_i}$, $i = 1, 2, 3, \cdots$, in (3.3) and (3.4), where $\mu_1, \mu_2, \mu_3, \cdots$ are any integers $\geqq 0$ such that if we let

$$\lambda_i = 2^{i-1}\lambda_1 + \mu_i$$

then $\lambda_{i+1} \geqq 2\lambda_i$, $i = 1, 2, 3, \cdots$. The coefficients in the continued fraction calculated from (3.5) are of the form $\pm c_1 c_{i+1} x^{\mu_{i+1}+1-2\mu_i}/c_i^2$, where $\mu_{i+1} - 2\mu_i = \lambda_{i+1} - 2\lambda_i \geqq 0$. Hence the continued fraction is of corresponding type; and it is readily seen to be the corresponding continued fraction for the resulting series, which is of the form (3.1). We thus find that the following theorem is true:

THEOREM 3.2. *The corresponding continued fraction for the series* (3.1) *in which* (3.2) *holds has the form*

$$c_0 + a_1 x^{\alpha_1}/1 + a_2 x^{\alpha_2}/1 + a_3 x^{\alpha_3}/1 + \cdots \tag{3.6}$$

*where the* $a_i$ *are independent of the* $\lambda_i$ *and are given by the formulas* (3.5); *and the* $\alpha_i$ *are independent of the* $c_i$ *and are given by the formulas*

$$\begin{aligned} &\alpha_1 = \lambda_1, \qquad \alpha_{2n} = \alpha_{2n+1}; \\ &\alpha_{2^{i+1}} = \lambda_1 + \lambda_{i+1} - 2\lambda_i, \quad \alpha_{n\cdot 2^{i+1}+2^{i+1}} = \alpha_{2^{i+1}}, \end{aligned} \qquad i, n = 1, 2, 3, \cdots. \tag{3.7}$$

The second proof of Theorem 3.2 which we shall now give is suggested by the form of the equations (3.5), (3.7), and will bring to light remarkable properties of the class of continued fractions under consideration.

We shall prove Theorem 3.2 for the case that the numbers $|c_i|$ are bounded and bounded away from 0, and the general case then readily follows. Under this restriction it will be seen from (3.5) that (3.6) converges uniformly in that vicinity of the origin which is determined by the inequalities $|a_n x^{\alpha_n}| \leqq \frac{1}{4}$, $n = 2, 3, 4, \cdots$ [Perron, p. 262]. Then by a theorem of Leighton and Scott [Leighton and Scott (1) p. 600] the power series corresponding to (3.6) converges in the vicinity of the origin to the analytic function $f(x)$ represented by this continued fraction. Our theorem will follow if we show that $f(x) = T(x)$, where $T(x)$ is the power series (3.1).

In order to do this, let $f = \left[c_0, \frac{a_n x^{\alpha_n}}{1}\right]$ be the continued fraction (3.6) and define the operator "$O$" by the relations:

$$Of = \left[c_0 + a_1 x^{\alpha_1}, \frac{-a_{2n-1}a_{2n}x^{\alpha_{2n-1}+\alpha_{2n}}}{1 + a_{2n}x^{\alpha_{2n}} + a_{2n+1}x^{\alpha_{2n+1}}}\right],$$

$$O^k f = O(O^{k-1}f), \quad k = 1, 2, 3, \cdots, \quad O^0 f = f.$$

Thus $Of$ is the continued fraction having as sequence of convergents the sequence of odd indexed convergents of (3.6) [Perron, p. 201]. Hence if $f$ converges so will $O^k f$, $k = 1, 2, 3, \cdots$, and we shall have $f = O^k f = f(x)$.

But by (3.5), (3.7) we have:

$$a_{2n}x^{\alpha_{2n}} = -a_{2n+1}x^{\alpha_{2n+1}}, \qquad a_{4n-1}x^{\alpha_{4n-1}} = -a_{4n+3}x^{\alpha_{4n+3}}, \cdots,$$

and therefore

$$O^k f = \left[c_0 + c_1 x^{\lambda_1} + c_2 x^{\lambda_2} + \cdots + c_k x^{\lambda_k}, \frac{a_n^{(k)} x^{\alpha_n^{(k)}}}{1}\right]$$

where

$$\begin{aligned} &\alpha_n^{(o)} = \alpha_n, \ a_n^{(o)} = a_n, \\ &\alpha_n^{(i)} = \alpha_{2n-1}^{(i-1)} + \alpha_{2n}^{(i-1)}, \qquad i, n = 1, 2, 3, \cdots. \\ &a_n^{(i)} = -a_{2n-1}^{(i-1)} a_{2n}^{(i-1)}, \end{aligned}$$

Thus we may write

$$f(x) = c_0 + c_1 x^{\lambda_1} + \cdots + c_k x^{\lambda_k} + x^{\lambda_{k+1}} g_k(x),$$

where $g_k(x)$ is analytic at $x = 0$. Consequently $f(0) = c_0$, $f^{(\lambda_n)}(0) = (\lambda_n)! c_n$, $n = 1, 2, 3, \cdots$, so that $f(x) = T(x)$, as was to be proved.

From the fact that the series $c_0 + c_{k+1}x^{\lambda_{k+1}} + c_{k+2}x^{\lambda_{k+2}} + \cdots$ has the form of the series $T(x)$, i.e. each exponent is at least twice the preceding, we see from the above proof that $a_n^{(k)}$ is the same function of $c_{k+1}, c_{k+2}, c_{k+3}, \cdots$ that $a_n$ is of $c_1, c_2, c_3, \cdots$; and $\alpha_n^{(k)}$ is the same function of $\lambda_{k+1}, \lambda_{k+2}, \lambda_{k+3}, \cdots$ that $\alpha_n$ is of $\lambda_1, \lambda_2, \lambda_3, \cdots$.

If $\lambda_1 = 1$ it is easy to verify that the inequalities (2.9) are satisfied by the $\alpha_n$ given by (3.7), and hence $T(x)$ is absolutely regular. Also, by the preceding

remark, the series $c_0 + c_{k+1}x + c_{k+2}x^{1+\lambda_{k+2}-\lambda_{k+1}} + c_{k+3}x^{1+\lambda_{k+3}-\lambda_{k+1}} + \cdots$ is absolutely regular. Hence if we remember that the convergents of $O^k f$ are among the convergents of $f$ we may apply Theorem 2.5 and have this result:

THEOREM 3.3. *If $\lambda_1 = 1$ the series $T(x)$ given by* (3.1), (3.2) *is absolute regular; and the convergents of the corresponding continued fraction fill all the fields $(i, j)$, $j \geqq i$, in the Padé table for $T(x)$.*

To facilitate the statement of subsequent theorems we shall introduce the notation

$$a_n = r_n(c_1, c_2, c_3, \cdots), \qquad \alpha_n = \rho_n(\lambda_1, \lambda_2, \lambda_3, \cdots)$$

to indicate the dependence of the $a_n$, $\alpha_n$ of (3.6) upon the numbers $c_i$, $\lambda_i$. In particular we shall put

$$\epsilon_n = r_n(1, 1, 1, \cdots). \tag{3.8}$$

THEOREM 3.4. *Let $T(x) = c_0 + \sum c_i x^{\lambda_i}$, $S(x) = d_0 + \sum d_i x^{\mu_i}$ be two power series in which each exponent is at least twice the preceding $(c_i \neq 0, d_i \neq 0)$, with corresponding continued fractions*

$$\left[c_0, \frac{\epsilon_n g_n x^{\alpha_n}}{1}\right], \qquad \left[d_0, \frac{\epsilon_n h_n x^{\beta_n}}{1}\right],$$

*respectively, where $\epsilon_n = \pm 1$ is given by* (3.8). *Then the "composition" series $c_0 d_0 + \sum c_i d_i x^{\lambda_i + \mu_i}$ has the corresponding continued fraction*

$$\left[c_0 d_0, \frac{\epsilon_n g_n h_n x^{\alpha_n + \beta_n}}{1}\right].$$

This theorem follows at once from the formulas (3.5), (3.7).

THEOREM 3.5. *Put $b_n = r_n(\lambda_2 c_2, \lambda_3 c_3, \lambda_4 c_4, \cdots)$, $n = 1, 2, 3, \cdots$. Then if $\lambda_1 = 1$ the power series obtained from* (3.1) *by differentiation term by term has the corresponding continued fraction*

$$c_1 + b_1 x^{\alpha_2}/1 + b_2 x^{\alpha_3 + \alpha_4}/1 + b_3 x^{\alpha_5 + \alpha_6}/1 + \cdots, \tag{3.9}$$

*where the $\alpha_n$ are given by* (3.7).

PROOF. The series $1 + xT'(x)$ is a series to which Theorem 3.2 can be applied. If the constant term and the factor $x$ be removed from this series and from its corresponding continued fraction, then the continued fraction can be thrown into the form (3.9) by contraction (the operation "$O$"), and since (3.9) is of corresponding type it must be the corresponding continued fraction for $T'(x)$.

THEOREM 3.6. *Let $\lambda_1 = 1$, $\lambda_{i+1} > 2\lambda_i$, $i = 1, 2, 3, \cdots$. Then the reciprocal of the series $T(x) = c_0 + \sum c_i x^{\lambda_i}$ $(c_i \neq 0)$ is absolutely regular and its corresponding continued fraction is*

$$c_0^{-1} + e_1 c_0^{-1} x^{\beta_1}/1 + e_2 x^{\beta_2}/1 + e_3 x^{\beta_3}/1 + \cdots \tag{3.10}$$

*where*

$$e_n = r_{n+1}(c_0, c_1, c_2, \cdots), \qquad \beta_n = \rho_{n+1}(1, 1 + \lambda_1, 1 + \lambda_2, \cdots), \qquad n = 1, 2, 3, \cdots.$$

PROOF. Since $\lambda_{i+1} > 2\lambda_i$, it follows that the series $1 + xT(x)$ is one in which each exponent is at least twice the preceding. Hence

$$1 + xT(x) \backsim 1 + c_0x/1 + e_1x^{\beta_1}/1 + e_2x^{\beta_2}/1 + \cdots,$$

and therefore $1/T(x)$ corresponds to (3.10). The absolute regularity of $1/T(x)$ may be proved by showing that the $\beta_n$ satisfy the inequalities (2.9).

When the conditions $\lambda_{i+1} > 2\lambda_i$, $i = 1, 2, 3, \cdots$, are not satisfied the continued fraction for the reciprocal of $T(x)$ does not have the property that the exponents in the partial numerators are independent of the coefficients $c_i$ in the power series. In the special case of the series

$$1 + x + x^2 + x^4 + x^8 + \cdots \tag{3.11}$$

we find the corresponding continued fraction for the reciprocal series to be

$$1 + d_1x/1 + d_2x^2/1 + d_3x/1 + d_4x/1 + \cdots, \tag{3.12}$$

where $d_1 = -1$, $d_2 = 1$, $d_3 = 2$; $d_{2n} + d_{2n+1} = \epsilon_{2n+1} + \epsilon_{2n+2}$, $d_{2n-1}d_{2n} = -1$, $n = 2, 3, 4, \cdots$, the $\epsilon_n$ being given by (3.8). It will be observed that (3.12) is not regular: $A_2/B_2$ is not a Padé approximant for (3.11).

A possible application of Theorem 3.6 is evident from the observation that *any neighborhood of the origin in which* (3.10) *is uniformly convergent is necessarily a zero-free region for the function* $T(x)$. Accordingly, we obtain from the Pringsheim criterion for uniform convergence [Perron, p. 262] this theorem:

THEOREM 3.7. *Let* $T(x)$ *be the power series of Theorem* 3.6 *and put* $q_i = \lambda_{i+1} - 2\lambda_i$, $i = 1, 2, 3, \cdots$, $q_0 = 1$,

$$B = \underset{0 \leqq n < +\infty}{\text{g.l.b.}} \left[ \frac{|c_n^2|}{4\,|c_0 c_{n+1}|} \right]^{\frac{1}{q_n}}. \tag{3.13}$$

*Then* $T(x) \neq 0$ *if* $|x| \leqq B$.

This result can be proved directly from the series $T(x)$, and indeed with the constant 4 in (3.13) replaced by a better constant $1/r$ where $r$ is at least .6. For if $B_r$ denotes the value of (3.13) with 4 replaced by $1/r$, then if $|x| \leqq B_r$,

$$|c_1x^{\lambda_1} + c_2x^{\lambda_2} + c_3x^{\lambda_3} + \cdots| \leqq \frac{|c_0|}{r} \sum_1^\infty r^{2^n},$$

and this will be less than $|c_0|$ if $r = .6$.

## 4. Absolutely regular continued fractions and power series

Let

$$\varphi_n = \begin{vmatrix} c_1, c_2, & \cdots, c_n \\ c_2, c_3, & \cdots, c_{n+1} \\ \cdot & \cdot & \cdot \\ c_n, c_{n+1}, & \cdots, c_{2n-1} \end{vmatrix}, \qquad \psi_n = \begin{vmatrix} c_2, c_3, & \cdots, c_n \\ c_3, c_4, & \cdots, c_{n+1} \\ \cdot & \cdot & \cdot \\ c_n, c_{n+1}, & \cdots, c_{2n-2} \end{vmatrix} \tag{4.1}$$

be the familiar persymmetric determinants formed from the coefficients of the power series $P(x) = 1 + \sum c_ix^i$. When these determinants are all different

from zero, and only then, $P(x)$ is *semi-normal*, i.e. in the corresponding continued fraction (1.1) for $P(x)$ the exponents $\alpha_n$ are all unity. Moreover, in this case the coefficients $a_n$ in (1.1) can be expressed in a simple way in terms of the determinants (4.1). We shall show how this result may be extended to all absolutely regular power series, (see Definition 2.2).

THEOREM 4.1. *Let* $P(x) = 1 + \sum_{i=1}^{\infty} c_i x^i$ *be an absolutely regular power series with corresponding continued fraction* (1.1). *Put*

$$g_n = \alpha_2 + \alpha_4 + \cdots + \alpha_{2n}, \qquad h_{n-1} = \alpha_1 + \alpha_3 + \cdots + \alpha_{2n-1},$$

$n = 1, 2, 3, \cdots$. *Then the determinants*

$$\varphi_{g_n}, \varphi_{h_n}, \psi_{1+g_n}, \psi_{h_n}, \qquad n = 1, 2, 3, \cdots, \tag{4.2}$$

*are all different from* 0, *and the coefficients* $a_n$ *in the corresponding continued fraction* (1.1) *are given by the formulas*:

$$\begin{gathered} a_1 = \varphi_{h_0}, \quad a_2 = (-1)^{\alpha_2}\psi_{1+g_1}/\varphi_{g_1}, \quad a_3 = (-1)^{\alpha_3}\varphi_{h_1}/\psi_{h_1}; \\ a_{2n} = (-1)^{\alpha_{2n}}\frac{\psi_{1+g_n}\varphi_{g_{n-1}}}{\varphi_{g_n}\psi_{1+g_{n-1}}}, \qquad a_{2n+1} = (-1)^{\alpha_{2n+1}}\frac{\varphi_{h_n}\psi_{h_{n-1}}}{\psi_{h_n}\varphi_{h_{n-1}}}, \quad n = 2, 3, 4, \cdots. \end{gathered} \tag{4.3}$$

PROOF. By (2.9) we know that $A_n(x)$ and $B_n(x)$ are polynomials of the form:

$$A_{2n}(x) = 1 + \sum_1^{g_n} p_i x^i, \qquad B_{2n}(x) = 1 + \sum_1^{g_n} q_i x^i,$$

$$A_{2n+1}(x) = 1 + \sum_1^{h_n} p_i' x^i, \qquad B_{2n+1}(x) = 1 + \sum_1^{h_n-1} q_i' x^i,$$

and we find by mathematical induction that

$$q_{g_n} = a_2 a_4 a_6 \cdots a_{2n}, \quad p'_{h_n-1} = a_1 a_3 a_5 \cdots a_{2n-1}, \qquad n = 1, 2, 3, \cdots. \tag{4.4}$$

By (1.4) the coefficients $p_i$, $q_i$, $p_i'$, $q_i'$, satisfy the following equations:

$$c_{g_n+j} + \sum_{i=1}^{g_n} q_i c_{g_n+j-i} = 0, \qquad j = 1, 2, \cdots, g_n; \tag{4.5}$$

$$\sum_{i=0}^{j} q_i c_{j-i} = p_j, \quad j = 1, 2, \cdots, g_n, q_0 = c_0 = 1; \tag{4.6}$$

$$c_{h_n+j} + \sum_{i=1}^{h_n-1} q_i' c_{h_n+j-i} = 0, \qquad j = 1, 2, \cdots, h_n; \tag{4.7}$$

$$\sum_{i=0}^{j} q_i' c_{j-i} = p_j', \qquad j = 1, 2, \cdots, h_n, q_0' = 1. \tag{4.8}$$

On account of the irreducibility of the fraction $A_n(x)/B_n(x)$ it follows that these equations have a unique solution; and consequently we find that the determinants $\varphi_{g_n}$, $\psi_{1+g_n}$, $\psi_{h_n}$, $\varphi_{h_n}$, must be different from zero. The formulas (4.3) then follow from (4.4).

Put

$$G_0^{(r)} = x^r,$$
$$G_1^{(r)} = x^{r-1} + c_1 x^r,$$
$$G_2^{(r)} = x^{r-2} + c_1 x^{r-1} + c_2 x^r,$$
$$\cdots\cdots\cdots\cdots$$
$$G_r^{(r)} = 1 + c_1 x + c_2 x^2 + \cdots + c_r x^r.$$

Then we have from (4.5) $\cdots$ (4.8) the following determinant expressions for the polynomials $A_n(x)$, $B_n(x)$:

$$A_{2n}(x) = \frac{(-1)^{g_n}}{\varphi_{g_n}} \begin{vmatrix} G_0^{(g_n)}, & G_1^{(g_n)}, & \cdots, & G_{g_n}^{(g_n)} \\ c_1, & c_2, & \cdots, & c_{1+g_n} \\ c_2, & c_3, & \cdots, & c_{2+g_n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{g_n}, & c_{1+g_n}, & \cdots, & c_{2g_n} \end{vmatrix},$$

(4.9)

$$B_{2n}(x) = \frac{(-1)^{g_n}}{\varphi_{g_n}} \begin{vmatrix} x^{g_n}, & x^{g_n-1}, & \cdots, & 1 \\ c_1, & c_2, & \cdots, & c_{1+g_n} \\ c_2, & c_3, & \cdots, & c_{2+g_n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{g_n}, & c_{1+g_n}, & \cdots, & c_{2g_n} \end{vmatrix},$$

$$A_{2n+1}(x) = \frac{(-1)^{h_n-1}}{\psi_{h_n}} \begin{vmatrix} G_1^{(h_n)}, & G_2^{(h_n)}, & \cdots, & G_{h_n}^{(h_n)} \\ c_2, & c_3, & \cdots, & c_{1+h_n} \\ c_3, & c_4, & \cdots, & c_{2+h_n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{h_n}, & c_{1+h_n}, & \cdots, & c_{2h_n} \end{vmatrix},$$

(4.10)

$$B_{2n+1}(x) = \frac{(-1)^{h_n-1}}{\psi_{h_n}} \begin{vmatrix} x^{h_n-1}, & x^{h_n-2}, & \cdots, & 1 \\ c_2, & c_3, & \cdots, & c_{1+h_n} \\ c_3, & c_4, & \cdots, & c_{2+h_n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{h_n}, & c_{1+h_n}, & \cdots, & c_{2h_n} \end{vmatrix}.$$

The integers $g_n$, $h_n$ satisfy the inequalities:

(4.11) $\quad 1 = h_0 \leqq g_1 \leqq h_1 - 1 \leqq g_2 - 1 \leqq h_2 - 2 \leqq g_3 - 2 \leqq h_3 - 3 \leqq \cdots.$

We now consider the converse of the last theorem. Given a set of integers $g_n$, $h_n$ satisfying (4.11) such that the determinants (4.2) are different from zero, we may construct the polynomials (4.9), (4.10), and the rational fractions

$A_n(x)/B_n(x)$ thus obtained will be Padé approximants for $P(x)$. Moreover, we may determine positive integers $\alpha_n$ from the relations

$$\alpha_1 = h_0 = 1, \quad \alpha_{2n+1} = h_n - h_{n-1}, \quad \alpha_2 = g_1, \quad \alpha_{2n} = g_n - g_{n-1}, \quad n = 1, 2, 3, \cdots \tag{4.12}$$

and numbers $a_n$ from (4.3) and hence form a continued fraction (1.1). However, there is no assurance that this will be the corresponding continued fraction for $P(x)$, nor, what is the same thing, that $A_n(x)/B_n(x)$ will be its $n^{\text{th}}$ convergent. We shall prove the following theorem:

THEOREM 4.2. *A power series* $P(x) = 1 + \sum_{i=1}^{\infty} c_i x^i$ $(c_1 \neq 0)$ *is absolutely regular if and only if there exist integers* $g_n$, $h_n$ *satisfying the inequalities* (4.11) *such that the determinants* (4.2) *are different from* 0, *and such that the polynomials* (4.9), (4.10) *satisfy an identity of the form*

$$A_n(x)B_{n-2}(x) - A_{n-2}(x)B_n(x) = H_n x^{\alpha_1+\alpha_2+\cdots+\alpha_{n+1}}, \tag{4.13}$$

$n = 2, 3, 4, \cdots$, *where* $H_n$ *is a constant* $\neq 0$, *and* $\alpha_1, \alpha_2, \alpha_3, \cdots$ *are given by* (4.12).

PROOF. The necessity of the condition follows from Theorem 4.1. The sufficiency of the condition will be established if we show that $A_n(x)/B_n(x)$ given by (4.9), (4.10) is the $n^{\text{th}}$ convergent of (1.1). From (4.13) it follows that

$$P(x)B_n(x) - A_n(x) = (x^{\alpha_1+\alpha_2+\cdots+\alpha_{n+1}})$$

and hence

$$A_n(x)B_{n+1}(x) - A_{n+1}(x)B_n(x) = K_n x^{\alpha_1+\alpha_2+\cdots+\alpha_{n+1}}$$

where $K_n$ is a constant $\neq 0$. Now let

$$y_0 + x_1/y_1 + x_2/y_2 + \cdots \tag{4.14}$$

be the continued fraction having $A_n(x)/B_n(x)$ for its $n^{\text{th}}$ convergent. Then we find that

$$y_0 = 1, \quad x_1 = c_1 x, \quad y_1 = 1,$$

$$x_n = -\frac{B_{n-2}(x)K_{n-1}x^{\alpha_n}}{B_n(x)K_{n-2}}, \quad y_n = -\frac{B_{n-1}(x)H_n}{B_n(x)K_{n-2}}, \quad n = 2, 3, \cdots,$$

from which it follows that (4.14) is a continued fraction of corresponding type.

## 5. Analytic functions having the unit circle as natural boundary

In this section we shall consider functions defined by corresponding type continued fractions of the form

$$1 + \frac{a_1 x^{n_1}}{1} + \cdots + \frac{a_k x^{n_k}}{1} + \frac{a_1 x^{n_1 m}}{1} + \cdots + \frac{a_k x^{n_k m}}{1} + \frac{a_1 x^{n_1 m^2}}{1} + \cdots, \tag{5.1}$$

in which $a_1, a_2, a_3, \cdots, a_k$ $(k \geqq 1)$ are real or complex numbers different from zero; $n_1, n_2, n_3, \cdots, n_k$ are positive integers; and $m$ is an integer $> 1$.

This continued fraction represents a function $f(x)$ which is meromorphic in any region $R$ whose closure lies entirely within the unit circle $|x| = 1$, and converges uniformly to $f(x)$ in the region obtained from $R$ by removing the interiors of small circles with centers at the poles of $f(x)$. In the neighborhood of the origin (5.1) and its corresponding power series converge uniformly to $f(x)$ [Leighton and Scott (2), p. 600]. Outside the unit circle (5.1) diverges by oscillation, the sequences of even and of odd convergents representing separate functions.

The function $f(x)$ satisfies the functional equation:

$$f(x) = \frac{f(x^m)A_{k-1}(x) + a_k x^{n_k} A_{k-2}(x)}{f(x^m)B_{k-1}(x) + a_k x^{n_k} B_{k-2}(x)}, \tag{5.2}$$

where we agree to write $A_{-1}(x) = 1$, $B_{-1}(x) = 0$ in case $k = 1$. Moreover, $f(x)$ is the only function analytic at $x = 0$ which satisfies this functional equation. This is a consequence of the one to one correspondence between power series and corresponding type continued fractions. We shall prove the following theorem.

THEOREM 5. *If there is a single point $x_0 = e^{it}$ upon the circle $|x| = 1$ where $f(x)$ is analytic, then $f(x)$ is a meromorphic function of $x$.*

PROOF. From (5.2) we have

$$f(z) = -\frac{A_{k-2}(x) - f(x)B_{k-2}(x)}{A_{k-1}(x) - f(x)B_{k-1}(x)} a_k x^{n_k}, \qquad z = x^m, \tag{5.3}$$

where the denominator does not vanish identically inasmuch as $B_{k-1}(x) \not\equiv 0$ and $f(x)$ does not represent a rational function of $x$.

Let $U$ be a neighborhood of $x_0$ in which $f(x)$ is analytic except possibly for a finite number of poles. As $x$ ranges over $U$, $z = x^m$ ranges over a region $V$ containing an arc of the unit circle $m$ ($\geqq 2$) times as long as that contained in $U$. If $x$ traverses a circle about any point of $U$, $z$ traverses a closed curve in $V$ once and only once and returns to its initial value. Hence $f(z)$ is a single-valued function of $z$ over $V$. Now the right member of (5.3) is an analytic function of $x$ over $U$ except for at most a finite number of poles. If $b$ is a point of $U$ where $f(x)$ is analytic then $f(z)$ can be expanded about $b^m$ in a series of analytic functions of $z$ which is uniformly convergent in some neighborhood of $b^m$, so that $f(z)$ is an analytic function of $z$ at $b^m$. Since a pole of the right member of (5.3) cannot correspond to a branch point or essential singularity of $f(z)$ in $V$ it must correspond to a pole of $f(z)$. Hence $f(z)$ is analytic over $V$ except possibly for a finite number of poles.

We may now repeat the above argument starting with $V$ instead of with $U$, and obtain a region $W$ containing an arc of the unit circle $m^2$ times as long as that contained in $U$, over which $f(x)$ is analytic except possibly for a finite number of poles. After a finite number of repetitions of this argument we find

that $f(x)$ has the specified property in an annulus containing the circle $|x| = 1$, of the form

$$1 - d \leqq r \leqq 1 + d, \qquad x = re^{i\theta}, \qquad d > 0.$$

Now by an argument similar to the above we find that $f(x)$ has the specified property in an annulus with inner radius $(1 - d)^m$ an outer radius $(1 + d)^m$. On repeating this argument a finite number of times one will eventually obtain an annulus overlapping the boundary of a neighborhood of the origin in which $f(x)$ is analytic, and extending out beyond any assigned circle $|x| = R$. Hence $f(x)$ is a meromorphic function.

THEOREM 5.2. *Let $s = 1$ or $-1$ if the exponents in (5.1) are all odd, and $s = 1$ if any of them are even. Put*

$$Z(s) = [B_k(s) - B_{k-1}(s) + A_{k-1}(s)]^2 - 4(-s)^k a_1 a_2 \cdots a_k .$$

*Then $f(x)$ has the unit circle as natural boundary if the $a_i$ are real, $B_{k-1}(s) \neq 0$, and $Z(s) < 0$.*

PROOF. Since the $a_i$ are real, $f(x)$ is real if $x$ is real and $-1 < x < 1$. If $s$ is a pole of $f(x)$, then (5.3) must be satisfied in some neighborhood of $x = s$, which would imply that $B_{k-1}(s) = 0$, contrary to hypothesis. If $f(x)$ is analytic at $x = s$ then $f(s)$ is a root of a quadratic equation with real coefficients and discriminant $Z(s) < 0$. But this is impossible inasmuch as $f(x)$ is real at some point in every neighborhood of $s$. Hence $s$ is a singularity of $f(x)$ which is not a pole. The theorem now follows by Theorem 5.1.

The case $|a_i| \leqq \frac{1}{4}$ is of particular interest because of the fact that in this case the continued fraction converges uniformly for $|x| \leqq 1$, and $f(x)$ is therefore analytic for $|x| < 1$ and continuous for $|x| \leqq 1$ [Perron (1), p. 262]. Hence if the corresponding power series has radius of convergence 1 we can conclude that $f(x)$ has a non-polar singularity on the circle $|x| = 1$, and must therefore have the unit circle as natural boundary by Theorem 5.1. We shall make use of this remark to prove the next theorem.

THEOREM 5.3. *If $a$ is real, the function*

$$f(x) = 1 + ax/1 + ax^m/1 + ax^{m^2}/1 + \cdots, \qquad m > 1,$$

*has the unit circle as natural boundary if*

(i) *$m$ is odd and $a \neq 0$;*

(ii) *$m$ is even and $a < 0$.*

*If $|a| \leqq \frac{1}{4}$, $f(x)$ is analytic for $|x| < 1$ and continuous for $|x| \leqq 1$.*

PROOF. The only case not covered by Theorem 5.2 with $k = 1$ is the case $|a| \leqq \frac{1}{4}$. We shall suppose $-\frac{1}{4} \leqq a < 0$. When $m$ is odd the case $0 < a \leqq \frac{1}{4}$ may be reduced to this by replacing $x$ by $-x$. Let $P(x)$ be the corresponding power series. It is required to show that the radius of convergence of $P(x)$ is 1. To do this give to $x$ a real fixed value $> 1$. Then the continued fraction may be considered as a Stieltjes continued fraction as a function of $a$. Let $H(a) = 1 + C_0 a - C_1 a^2 + C_2 a^3 - \cdots$ be the corresponding Stieltjes series.

The coefficients $C_i$ are polynomials in $x$ with positive integral coefficients [Stieltjes (3), p. 420]. Moreover, $H(a)$ is divergent except when $a = 0$, [Stieltjes (3), p. 424].

Let $C_j = \sum_{i=0}^{n_j} b_{ij}x^i$. Then

$$H(a) = 1 + \sum_{j=0}^{\infty}(-1)^j C_j a^{j+1} = 1 - \sum_{j=0}^{\infty}\sum_{i=0}^{n_j} b_{ij} \mid a \mid^{j+1} x^i,$$

since $a$ is real and negative. Now if $P(x)$ were convergent, it would follow that $H(a)$ would be convergent for values of $a$ other than 0. This contradiction shows that the radius of convergence of $P(x)$ is not greater than 1. Neither is it less than 1 inasmuch as the continued fraction converges uniformly for $\mid x \mid \leqq 1$.

It should be remarked that the power series for the function $f(x)$ of Theorem 5.3 does not in general have Hadamard gaps. For if in the functional relation

$$f(x) = 1 + ax/f(x^m)$$

we put

$$f(x) = 1 + \sum_{i=1}^{\infty} c_i x^i$$

there results the power series identity

$$\left(\sum_{i=1}^{\infty} c_i x^i\right)\left(1 + \sum_{i=1}^{\infty} c_i x^{mi}\right) = ax.$$

The coefficient of $x^{mi+j}$, $j = 1, 2, \cdots, m$, is

$$c_{mi+j} + \sum_{s=1}^{i} c_s c_{m(i-s)+j} = 0,$$

except for $i = 0, j = 1$, in which case $c_1 = a$. It can easily be shown that

$$c_{mi+j} = 0, \qquad j = 2, 3, \cdots, m.$$

Thus the $c_s$ appearing in the summation is zero unless $s = mt + 1$, in which case we can write

$$c_{m(mp+q)+1} + \sum_{t=0}^{p} c_{mt+1} c_{m[m(p-t)+q-1]+1} = 0,$$

where $q = 1, 2, \cdots, m$. This last relation can be used to show that $c_{m(mp+q)+1}$ is a polynomial in $a$ which is not $\equiv 0$.

When $k = 2$ in Theorem 5.2 we have

THEOREM 5.4. *Let $u$, $v$ be parameters, $u \neq v$, and $m$ an integer $> 1$. Then the continued fraction*

$$1 - (u+v)x/1 - (u-v)x/1 - (u+v)x^m/1 - (u-v)x^m/1 - (u+v)x^{m^2}/1 - \cdots \tag{5.4}$$

*represents a function $f(x)$ having the unit circle as natural boundary if $u$, $v$ are real and $u > v^2 + \frac{1}{4}$. If $m$ is odd the same holds if $u < -v^2 - \frac{1}{4}$.*

When $u = 0$ in (5.4) we find by contraction and application of Theorem 5.3 that $f(x)$ has the unit circle as natural boundary if $v$ is pure imaginary, and $m$ is even or odd.

For $k = 3, 4$, the conditions for a natural boundary of Theorem 5.2 are:

$$[1 + (a_1 + a_2 + a_3)s]^2 + 4sa_1a_2a_3 < 0, \qquad 1 + a_2s \neq 0;$$

and

$$[1 + (a_1 + a_2 + a_3 + a_4)s + (a_1a_3 + a_2a_4)]^2 - 4a_1a_2a_3a_4 < 0, \; 1 + (a_2 + a_3)s \neq 0,$$

respectively.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILLINOIS.

## REFERENCES

(1) Perron: *Die Lehre von den Kettenbrüchen*, Leipzig und Berlin (1929), Second Edition.
(2) Leighton and Scott: Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 596–605.
(3) Stieltjes: *Oeuvres* vol. 2.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# ON ZEROS OF IRREGULAR TAYLOR'S SERIES, AND AN APPROXIMATION PROBLEM

BY WALTER STRODT

(Received May 29, 1939)

**1. Introduction.** Let $f(z)$ be an integral function. If $\rho$, $r$, $\sigma$ are positive numbers such that $\rho < r$, and $\sigma > 1$, then we shall say that the triple $(\rho, r, \sigma)$ is a *gap-index* of $f(z)$ if $f(z)$ can be written in the form

$$f(z) = \sum_{n=0}^{\infty} b_{\lambda_n} z^{\lambda_n} + \sum_{n=0}^{\infty} b_{\mu_n} z^{\mu_n} \tag{1}$$

where,

(2) the $\lambda$'s and $\mu$'s are non-negative integers, with $\lambda_{n+1} > \lambda_n$, $\mu_{n+1} > \mu_n$,

(3) the function $\sum_{n=0}^{\infty} b_{\mu_n} z^{\mu_n}$ is of order not greater than $\rho$,

(4) the function $\sum_{n=0}^{\infty} b_{\lambda_n} z^{\lambda_n}$ is of order not greater than $r$, and

(5) for every positive $\epsilon$ the inequality $\lambda_{n+1} > \lambda_n^{\sigma-\epsilon}$ is valid for infinitely many values of $n$.

If $f(z)$ has gap-index $(\rho, r, \sigma)$, and is actually of order $r$, we shall say that $f(z)$ is a function of *irregular development.*

Maillet has proved that a function of irregular development is not of *regular growth,* in the sense of Borel.[1] Borel has shown[2] that if $f(z)$ is of finite positive order and is not of regular growth, and if $a_1, a_2, \cdots$, are the zeros of $f(z)$, then in the case where the limiting exponent of the zeros is equal to the order of $f(z)$ we have

$$\limsup_{n\to\infty} \frac{\log |a_n|}{\log n} > \liminf_{n\to\infty} \frac{\log |a_n|}{\log n}. \tag{6}$$

From these two results we get the following corollary:

THEOREM A. *If $f(z)$ is a function of irregular development, with zeros $a_1, a_2, \cdots$, such that the limiting exponent of the zeros is equal to the order of $f(z)$, then the inequality* (6) *is valid.*

We shall prove directly a theorem which is somewhat more precise than Theorem A. Our method is based essentially on the possibility of approxi-

[1] E. Maillet, *Sur les fonctions entières,* Annales de la Faculte des Sciences de Toulouse, 2nd series, v. 4, (1902), p. 455. The converse has been proved by E. Lindelöf, *Sur la determination de la croissance des fonctions entières,* Bulletin des Sciences Mathématiques, v. 27, (1903), p. 219. A more precise statement is given by J. M. Whittaker, *The lower order of integral functions,* Journal of the London Mathematical Society, v. 8, (1933), p. 20.

[2] E. Borel, *Leçons sur les fonctions entières,* pp. 107, 108, (Gauthiers-Villars, 1900). Borel proves the converse as well.

mating an arbitrary integral function uniformly in every bounded region by functions of the form

$$\sum_{i=1}^{s} c_i f(a_i z), \tag{7}$$

where the $c$'s are arbitrary complex numbers. In connection with this approximation problem we obtain results supplementing those of Gelfond,[3] our methods moreover being entirely different from his.

**2.** We shall use the following lemma:

LEMMA 1. *Let $f(z)$ be an integral function with gap-index $(\rho, r, \sigma)$. Then the integral function $f(z)f(-z)$ has gap-index $(\rho, r, \sigma)$.*

PROOF: Let statements (1), (2), (3), (4), (5) be valid. Then

$$f(z)f(-z) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\{(-1)^{\lambda_m} b_{\lambda_n} b_{\lambda_m} z^{\lambda_n+\lambda_m} + (-1)^{\mu_m} b_{\mu_n} b_{\mu_m} z^{\mu_n+\mu_m} + [(-1)^{\mu_m} + (-1)^{\lambda_n}] b_{\lambda_n} b_{\mu_m} z^{\lambda_n+\mu_m}\}. \tag{8}$$

We write $f(z)f(-z) = g(z) + h(z)$, where in $g(z)$ are put all the terms from the right of (8) which have exponent $\mu_n + \mu_m$, and all which have exponent $\lambda_n + \mu_m$ with $\mu_m > \lambda_n \log \lambda_n$, and in $h(z)$ are put all the remaining terms from the right of (8). It is easy to see that $g(z)$ is of order not greater than $\rho$, while $h(z)$ is, of course, of order not greater than $r$; moreover, in $h(z)$, for every positive $\epsilon$, the smallest exponent of $z$ effectively appearing and exceeding $\lambda_p(1 + \log \lambda_p)$ is greater than $[\lambda_p(1 + \log \lambda_p)]^{\sigma-\epsilon}$ for infinitely many values of $p$. This evidently suffices to prove the lemma.

COROLLARY: *Let the even function $f(z)f(-z)$ be written $f(z)f(-z) = F(z^2)$. Then $F(z)$ has gap-index $(\rho/2, r/2, \sigma)$.*

**3.** Our principal theorem on zeros is the following:

THEOREM 1. *Let $f(z)$ be an integral function which is not identically zero, and which has gap-index $(\rho, r, \sigma)$. Let the zeros of $f(z)$ be written in a sequence, a multiple zero being written a number of times equal to its order of multiplicity. Let $a_1, a_2, \cdots$ be a subsequence of this sequence. Let*

$$l = \limsup_{n\to\infty} \frac{\log |a_n|}{\log n}.$$

*Then $l \geqq \min(1/\rho, \sigma/r)$.*

PROOF. We make a preliminary assumption that the $a$'s are distinct. Now suppose that $l < \min(1/\rho, \sigma/r)$. We consider first the following special case:

---

[3] A. Gelfond, *Sur les systemes complets de fonctions analytiques*, Recueil Mathématique, new series, v. 4(46): 1, (1938), p. 149.

CASE 1. There exist positive numbers $\tau$, $\sigma_0$, $\delta$ such that

$$l < 1/\tau < \min(1/\rho, \sigma/r), \qquad \rho + 2\delta < \tau, \tag{9}$$

$$r + 2\delta < \sigma_0\tau, \qquad \sigma_0 < \sigma, \tag{10}$$

$$(\rho + \delta)^{-1} - (\rho + 2\delta)^{-1} > 1, \qquad (r + \delta)^{-1} - (r + 2\delta)^{-1} > 1. \tag{11}$$

Of course, we can always find $\tau$, $\sigma_0$, $\delta$ to satisfy (9) and (10). It is condition (11) which makes Case 1 special.

We shall prove that for every complex number $\zeta$, every positive number $R$, and every positive number $\epsilon$, we can choose complex numbers $c_1, \cdots, c_s$ so that

$$\left| f(\zeta z) - \sum_{k=1}^{s} c_k f(a_k z) \right| < \epsilon, \tag{12}$$

for all $z$ with $|z| < R$. Now since $f(a_k) = 0$ for every $k$, this approximation considered at the point $z = 1$ will give us the contradiction that $f(\zeta) = 0$ for every complex $\zeta$. Finally we shall complete the proof of the theorem by freeing ourselves from the restrictions of being in Case 1 and of having distinct $a$'s.

Let $R$, $\epsilon$ be any positive numbers, $\zeta$ any complex number.

Let $f(z) = \sum_{n=0}^{\infty} b_n z^n$. Then for every set of complex numbers $c_1, \cdots, c_s$ we have

$$f(\zeta z) - \sum_{k=1}^{s} c_k f(a_k z) = \sum_{n=0}^{\infty} b_n \left( \zeta^n - \sum_{k=1}^{s} c_k a_k^n \right) z^n. \tag{13}$$

We assert that we can determine $c_1, \cdots, c_s$ so that the left-hand member of (13) has a zero of order $s$ at the origin, that is, so that

$$\zeta^n - \sum_{k=1}^{s} c_k a_k^n = 0, \qquad (n = 0, 1, \cdots, s-1). \tag{14}$$

This follows from the fact that the equations (14) are linear in the $c$'s with a determinant of coefficients which is equal to the Vandermondian of the $a$'s and which is therefore not zero. From now on we assume that the $c$'s are specialized to satisfy (14). Let

$$B_n = \sum_{k=1}^{s} c_k a_k^n, \qquad (n = 0, 1, \cdots). \tag{15}$$

Of course $\zeta^n - B_n = 0$ when $n < s$. We seek an upper bound for $\zeta^n - B_n$, when $n \geqq s$.

Let $p(x) = x^s + t_1 x^{s-1} + \cdots + t_s$ be that polynomial with unity for leading coefficient which has $a_1, a_2, \cdots, a_s$ for roots. If we take the equations (15) for $n = m, m+1, \cdots, m+s$, multiply the $j^{\text{th}}$ by $t_{s-j+1}$ $(j = 1, 2, \cdots, s)$, and add, we see that

$$B_{s+m} + t_1 B_{s+m-1} + \cdots + t_s B_m = \sum_{k=1}^{s} c_k a_k^m p(a_k) = 0 \qquad (m = 0, 1, \cdots). \tag{16}$$

Let $T_1 = |a_1| + |a_2| + \cdots + |a_s|$, $P = (|\zeta| + |a_1|) \cdots (|\zeta| + |a_s|)$. Let $s$ be so large that $|\zeta| < T_1$. We assert that

$$|\zeta^{s+n} - B_{s+n}| \leqq (3T_1)^n P, \qquad (n = 0, 1, \cdots). \tag{17}$$

This is easily proved by induction, with the use of the recursive relations (16). Now since

$$\limsup_{n\to\infty} \frac{\log |a_n|}{\log n} < \frac{1}{\tau},$$

there is a positive $C$ such that for all $n$

$$|\zeta| + |a_n| < Cn^{1/\tau}. \tag{18}$$

Hence $T_1 < Cs^{(1/\tau+1)}$, $P < C^s s^{s/\tau}$ for all $s$. Therefore, by (17),

$$|\zeta^{s+n} - B_{s+n}| < (3C)^{s+n} s^{s+n} s^{(s+n)/\tau}$$

for $s$ sufficiently large.

Now let statements (1), (2), (3), (4), (5) be satisfied. Then, if $|z| < R$,

$$\left|f(\zeta z) - \sum_{k=1}^{s} c_k f(a_k z)\right| < \sum_{\lambda_n \geqq s} |b_{\lambda_n}|\,|\zeta^{\lambda_n} - B_{\lambda_n}|\,R^{\lambda_n} + \sum_{\mu_n \geqq s} |b_{\mu_n}|\,|\zeta^{\mu_n} - B_{\mu_n}|\,R^{\mu_n}$$

$$< \sum_{\lambda_n \geqq s} \lambda_n^{-\frac{\lambda_n}{r+\delta}} A^{\lambda_n} s^{\lambda_n(1+1/\tau)} + \sum_{\mu_n \geqq s} \mu_n^{-\frac{\mu_n}{\rho+\delta}} A^{\mu_n} s^{\mu_n(1+1/\tau)},$$

(for $s$ sufficiently large, and assuming $A = 3CR$),

$$= \sum_{\lambda_n \geqq s} \left(s^{1/\tau}/\lambda_n^{\left(\frac{1}{r+2\delta}\right)}\right)^{\lambda_n} \left(sA/\lambda_n^{\left(\frac{1}{r+\delta} - \frac{1}{r+2\delta}\right)}\right)^{\lambda_n}$$

$$+ \sum_{\mu_n \geqq s} \left(s^{1/\tau}/\mu_n^{\left(\frac{1}{\rho+2\delta}\right)}\right)^{\mu_n} \left(sA/\mu_n^{\left(\frac{1}{\rho+\delta} - \frac{1}{\rho+2\delta}\right)}\right)^{\mu_n}.$$

Now take $s - 1 = \lambda_p$, where $\lambda_{p+1} > (1 + \lambda_p)^{\sigma_0}$. Then for every $\lambda_n$ such that $\lambda_n \geqq s$, we have $\lambda_n > s^{\sigma_0}$, so that $s^{1/\tau} < \lambda_n^{\left(\frac{1}{r+2\delta}\right)}$, by (10). Taking into account relations (9) and (11), we see now that we can choose $s$ to secure (12). Hence we reach the anticipated contradiction that $f(z) \equiv 0$. We now free ourselves from the restriction of having Case 1. We choose $\tau$, $\sigma_0$, $\delta$ so that relations (9) and (10) hold, and we set

$$u = (\rho + \delta)^{-1} - (\rho + 2\delta)^{-1} > 0$$

$$v = (r + \delta)^{-1} - (r + 2\delta)^{-1} > 0.$$

We now introduce new functions $\{f_n(z)\}$ as follows:

$$f_1(z) = f(z^{\frac{1}{2}}) f(-z^{\frac{1}{2}})$$

$$f_m(z) = f_{m-1}(z^{\frac{1}{2}}) f_{m-1}(-z^{\frac{1}{2}}), \qquad (m = 2, 3, \cdots).$$

We fix $m$ so that $2^m u > 1$, $2^m v > 1$, and we set $\alpha = 2^m$. Then $f_m(z)$ has the zeros $a_1^\alpha, a_2^\alpha, \cdots$. We define $\rho' = \rho/\alpha$, $r' = r/\alpha$, $\tau' = \tau/\alpha$, $\delta' = \delta/\alpha$, $l' = \alpha l$, $a_n' =$ the $n^{\text{th}}$ distinct number in the sequence $\{a_n^\alpha\}$. Then $f_m(z)$ has gap-index $(\rho', r', \sigma)$, by the corollary of Lemma 1, and

$$l' = \limsup_{n\to\infty} \frac{\log |a_n'|}{\log n}$$

and all the relations

$$(9') \qquad l' < 1/\tau' < \min\,(1/\rho', \sigma/r'), \qquad \rho' + 2\delta' < \tau',$$

$$(10') \qquad r' + 2\delta' < \sigma_0 \tau', \qquad \sigma_0 < \sigma,$$

$$(11') \quad (\rho' + \delta')^{-1} - (\rho' + 2\delta')^{-1} > 1, \qquad (r' + \delta')^{-1} - (r' + 2\delta')^{-1} > 1,$$

are valid. We can now prove that $f_m(z) = 0$, by the argument used above for $f(z)$. This contradiction gives us the theorem, for the case where the $a_k$ are distinct. The general case is treated in a similar fashion; for each number $a$ appearing $q$ times in the sequence $\{a_n\}$ we introduce the functions $f(az)$, $zf'(az), \cdots, z^{q-1}f^{(q-1)}(az)$,[4] and show that every function $f(\zeta z)$ can be approximated uniformly in every bounded region by finite linear combinations of functions in the set so obtained. We omit the remaining details.

**4.** Now if $f(z)$ is of finite order $r$, with zeros $a_1, a_2, \cdots$, and if the limiting exponent of the zeros is equal to the order of $f(z)$, then

$$(19) \qquad \liminf_{n\to\infty} \frac{\log |a_n|}{\log n} \leqq \frac{1}{r}.$$

If moreover $f(z)$ is of irregular development, then there exist numbers $k$ greater than 1 such that $(r/k, r, k)$ is a gap-index of $f(z)$. Let $\sigma$ be the supremum of such numbers $k$. From Theorem 1 and (19) we obtain

$$(20) \qquad \frac{\limsup_{n\to\infty} (\log |a_n|)/(\log n)}{\liminf_{n\to\infty} (\log |a_n|)/(\log n)} \geqq \sigma.$$

This is a more precise formulation of Theorem A.

**5.** The proof of Theorem 1 raises the following question: If $f(z)$, $g(z)$ are given integral functions, and if $a_1, a_2, \cdots$ are given (distinct) complex numbers, can $g(z)$ be approximated uniformly in every bounded region by functions of the form (7)? An affirmative answer is given by the proof above under the restrictions that $g(z)$ is in the form $f(\zeta z)$ and that we have Case 1. Using Theorem 1 we can now state the following more general result on approximation:

[4] Superscripts are indices of differentiation.

THEOREM 2. *If $f(z)$ has gap-index $(\rho, r, \sigma)$, and if*

$$\limsup_{n\to\infty} \frac{\log |a_n|}{\log n} < \min (1/\rho, \sigma/r),$$

*then every integral function $g(z)$ such that*

$$E\{n; g^{(n)}(0) \neq 0\} \subseteq E\{n; f^{(n)}(0) \neq 0\}$$

*can be approximated uniformly in every bounded region by functions of the form* (7).

PROOF: We note first that uniform convergence in $|z| < R'$ is implied by convergence in the $L_2(0, 2\pi)$ metric on $|z| = R > R'$ by the Cauchy integral theorem, and by Schwarz's inequality. We prove then that the closed linear manifold $\mathfrak{M}$ in $L_2(0, 2\pi)$ determined by the set of functions $\{e^{in\theta}; f^{(n)}(0) \neq 0\}$, is also determined by the set of functions $\{f(a_k R e^{i\theta}), k = 1, 2, \cdots\}$. This follows from the fact that if $\varphi(\theta)$ is an element of $\mathfrak{M}$ such that

$$\int_0^{2\pi} f(a_k R e^{i\theta}) \overline{\varphi(\theta)}\, d\theta = 0, \qquad (k = 1, 2, \cdots),$$

then $\varphi(\theta) = 0$; and this in turn follows[5] from the fact that the integral function $F(\zeta) \equiv \int_0^{2\pi} f(\zeta R e^{i\theta}) \overline{\varphi(\theta)}\, d\theta$ has gap-index $(\rho, r, \sigma)$, and so cannot vanish at $\zeta = a_k$, $(k = 1, 2, \cdots)$, without vanishing identically, (by Theorem 1).[6]

We observe finally that in the *general* case where the hypotheses of Theorem 2 are satisfied the functions (7), *chosen as in Case* 1, conceivably might fail to give uniform approximation to the function $f(\zeta z)$.

PRINCETON UNIVERSITY.

[5] The device used in this step was suggested to me by Professor S. Bochner.

[6] Identical vanishing of $F(\zeta)$ is easily seen to imply that $\varphi(\theta) = 0$.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# ON A PROPERTY OF $W$-CONGRUENCES

BY GUIDO FUBINI

(Received April 17, 1939)

In a recent paper (Math. Ann. vol. 114, p. 237), Jonas has given a very interesting new property of $W$-congruences. After some general remarks, this paper will give here new, easy, and purely projective demonstrations of these theorems. On the last page we state a new problem which seems to be important for the theory of $W$-congruences.

We shall say that two surfaces, whose points are in a given one-to-one reciprocal correspondence, are *congruence-transforms* of each other if they are the focal surfaces of the congruence generated by the straight lines joining two corresponding points of the two surfaces.

We shall denote by $x, y, \cdots$ the points whose homogeneous projective coordinates are $x_i, y_i, \cdots$ $(i = 1, 2, 3, 4)$, by $ax + by + \cdots$ the point whose coördinates are $ax_i + by_i + \cdots$. If $\varphi$ is a function of two parameters $u, v$, we denote by $\varphi_u, \varphi_v, \varphi_{uu}, \varphi_{uv}, \cdots$ the derivatives $\partial\varphi/\partial u$, $\partial\varphi/\partial v$, $\partial^2\varphi/\partial u^2$, $\partial^2\varphi/\partial u\partial v, \cdots$. When $x_i$ are functions of $u, v$, we denote by $x_u, x_v, \cdots$ the points whose coördinates are $\partial x_i/\partial u, \partial x_i/\partial v, \cdots$. By $(x, y, z, t)$ we denote the determinant

$$\begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ t_1 & t_2 & t_3 & t_4 \end{vmatrix}.$$

By $(x, y)$ we denote the matrix $\begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \end{vmatrix}$ and its minors, which are the coördinates of the straight line $(x, y)$ joining the points $x, y$.

We shall write also:

$$(x, y, z, t) = \sum (x, y)(z, t).$$

By the right-hand member we denote the sum

$$\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \cdot \begin{vmatrix} y_3 & y_4 \\ t_3 & t_4 \end{vmatrix} + \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \begin{vmatrix} z_4 & z_2 \\ t_4 & t_2 \end{vmatrix} + \cdots,$$

where every minor of $(x, y)$ is multiplied by its complement or complementary minor in the determinant $(x, y, z, t)$.

Let the $x_i$ and $y_i$ be functions of two parameters $u, v$. Let the surfaces $S_x$ and $S_y$, loci of the points $x(u, v)$ and $y(u, v)$, be in a one-to-one reciprocal corre-

spondence, and let corresponding points have the same curvilinear coördinates $u, v$. If we give the values of $u$, $v$ and of the slope $dv:du$, we determine on the surfaces $S_x$, $S_y$ two corresponding points $x$, $y$, and two corresponding tangents $t_x$, $t_y$. The first is the line joining the points $x$ and $dx = x_u du + x_v dv$; the second is the line passing through $y$, $dy$. If we change the slope into $\delta v:\delta u$, we obtain two other tangents $\tau_x$, $\tau_y$ joining $x$ with $\delta x = x_u \delta u + x_v \delta v$, and $y$ with $\delta y$. The *non*-corresponding tangents $t_x$ and $\tau_y$ will meet if and only if

$$0 = (x, y, dx, \delta y),$$

i.e., if

$$\begin{aligned} 0 = {} & (x, y, x_u, y_u)\, du\delta u + (x, y, x_u, y_v)\, du\delta v \\ & + (x, y, x_v, y_u)\, dv\delta u + (x, y, x_v, y_v)\, dv\delta v. \end{aligned} \tag{1}$$

This equation defines a collineation $T$ between the pencils of the lines $t_x$, $\tau_y$ (or between the slopes $dv:du$ and $\delta v:\delta u$). This collineation is an involution if and only if

$$(x, y, x_u, y_v) = (x, y, x_v, y_u). \tag{2}$$

$t_y$ and $\tau_x$ intersect provided (2) is satisfied, and provided $t_x$ and $\tau_y$ meet also. If the double elements of the collineation are distinct, by a suitable choice of the parameters $u$, $v$ we can obtain that they are the elements $du = 0$, $dv = 0$. With this choice of $u$, $v$ it is obvious that

$$(x, y, x_u, y_u) = (x, y, x_v, y_v) = 0, \tag{3}$$

and that $u = \text{const.}$, $v = \text{const.}$ will be the equations of the developables of the congruence generated by the lines $(x, y)$ joining two corresponding points $x$, $y$. If $X$, $Y$ are the foci[1] of this line, $X$, $Y$, $X_u$, $Y_v$ are points of the line $(x, y)$; therefore we can determine eight functions $l$, $m$, $\lambda$, $\mu$, $p$, $q$, $\pi$, $\kappa$ of $u$, $v$ such that

$$\begin{aligned} x &= lX + mY, & y &= \lambda X + \mu Y, \\ X_u &= pX + qY, & Y_v &= \pi X + \kappa Y. \end{aligned} \tag{4}$$

Obviously we can write the equation (2) in the form

$$(X, Y, x_u, y_v) = (X, Y, x_v, y_u).$$

By means of (4) this equation may be written

$$(X, Y, mY_u, \lambda X_v) = (X, Y, lX_v, \mu Y_u),$$

which becomes

$$m\lambda + l\mu = 0 \tag{5}$$

Therefore: *A necessary and sufficient condition that the collineation $T$ be an involution, is that the points $x$, $y$ be conjugate with respect to the foci of the line*

[1] We denote by $X$ the points of the edge of regression of a developable $v = \text{const.}$, and by $Y$ the points of the edge of regression of a developable $u = \text{const.}$

$(x, y)$. [The foci, which we have supposed distinct, are calculated with respect to the congruence generated by the lines $(x, y)$.]

We now suppose that $T$ is an involution, that $du = 0$ and $dv = 0$ are no longer its double elements, but that $du = 0$ and $\delta v = 0$ are two conjugate directions. In this case we can obviously deduce from (1) and (2) that

$$(6) \qquad (x, y, x_u, y_v) = (x, y, x_v, y_u) = 0.$$

Let $z$ be the intersection of the lines $(x, x_u)$ and $(y, y_v)$; and let $t$ be the intersection of the tangents $(x, x_v)$ and $(y, y_u)$. We can find eight functions $l, m, \cdots$ of $u, v$, such that

$$(7) \qquad \begin{aligned} x_u &= \lambda x + \mu z, \qquad & x_v &= lx + mt, \\ y_u &= ny + rt, \qquad & y_v &= \nu y + \rho z, \end{aligned}$$

[the foci of the line $(x, y)$ are the points $x\sqrt{r\rho} \pm y\sqrt{m\mu}$]. Is it possible that *the plane osculating at the point $x$ the curve $v =$ const. of the surface $S_x$ is identical with the plane osculating at the point $y$ the curve $u =$ const. on $S_y$?* In this case we deduce from (7) that both the planes will coincide with the plane $(x, y, z)$ passing through the points $x, y, z$. This is possible only if the points $x_{uu}$ and $y_{vv}$ lie on this plane. But from (7) we get:

$$x_{uu} = (\lambda_u + \lambda^2)x + (\mu_u + \lambda\mu)z + \mu z_u,$$

$$y_{vv} = (\nu_v + \nu^2)y + (\rho_v + \rho\nu)z + \rho z_v.$$

The points $x_{uu}, y_{vv}$ lie in the plane $(x, y, z)$ only if this plane passes through the points $z_u, z_v$; i.e., if we can find six functions $A, B, C, P, Q, R$ such that

$$(8) \qquad z_u = Ax + By + Cz, \qquad z_v = Px + Qy + Rz.$$

The geometrical interpretation of (8) is that *the plane $(x, y, z)$ envelopes the surface $S_z$* locus of the point $z$. Therefore *both the surfaces $S_x$, $S_y$ will be congruence-transforms of the same surface $S_z$.* In the same way we demonstrate that *the plane osculating at the point $x$ the curve $u =$ const. of the surface $S_x$ is identical with the plane osculating at the corresponding point $y$ the curve $v =$ const.* on $S_y$ only if it is possible to find six other functions $L, M, N, J, H, K$ such that

$$(9) \qquad t_u = Lx + My + Nt, \qquad t_v = Jx + Hy + Kt.$$

Only in this case are *both the surfaces $S_x$, $S_y$ congruence-transforms* of the same surface $S_t$ [and the plane $(x, y, t)$ envelopes the surface $S_t$]. The conditions of integrability of (7) are:

$$(10) \qquad \frac{\partial}{\partial v} x_u = \frac{\partial}{\partial u} x_v; \qquad \frac{\partial}{\partial v} y_u = \frac{\partial}{\partial u} y_v.$$

By expanding the first of these equations we find:

$$\lambda_v x + \mu_v z + \lambda(lx + mt) + \mu(Px + Qy + Rz)$$
$$= l_u x + m_u t + l(\lambda x + \mu z) + m(Lx + My + Nt).$$

This equation is satisfied if and only if

$$\begin{aligned} &\mu R + \mu_v = l\mu, \qquad mN + m_u = \lambda m, \\ &\lambda_v + \mu P = l_u + mL, \qquad \mu Q = mM. \end{aligned} \tag{11}$$

From the second of the (10) we deduce in the same way

$$rK + r_v = r\nu; \qquad \rho C + \rho_u = \rho n. \tag{12}$$

The minors $(x, y)$ of the matrix $(x, y)$ are the coördinates of the line joining the points $x, y$. From (7, 8, 9, 10, 11) we deduce

$$(x, y)_u = (x_u, y) + (x, y_u) = (\lambda + n)(x, y) + \mu(z, y) + r(x, t);$$

$$(x, y)_v = \qquad\qquad = (l + \nu)(x, y) + m(t, y) + \rho(x, z);$$

$$\begin{aligned} (x, y)_{uv} = \frac{\partial}{\partial v}(x, y)_u &= (\lambda + n)_v(x, y) + (\lambda + n)(x, y)_v + \mu[(z_v, y) + (z, y_v)] \\ &+ r[(x_v, t) + (x, t_v)] + \mu_v(z, y) + r_v(x, t) \\ &= \Lambda(x, y) + (\lambda + n)(x, y)_v + \mu_v(z, y) + \mu(R + \nu)(z, y) \\ &+ r_v(x, t) + r(l + K)(x, t) \\ &= \Lambda(x, y) + (\lambda + n)(x, y)_v + (l + \nu)[n(z, y) + r(x, t)] \\ &= P(x, y) + (\lambda + n)(x, y)_v + (l + \nu)(x, y)_u, \end{aligned}$$

where $\Lambda$ and $P$ are functions, whose values we shall not calculate. Hence the coördinates of the line $(x, y)$ satisfy the equation

$$\begin{aligned} &(x, y)_{uv} + \alpha(x, y)_u + \beta(x, y)_v + \gamma(x, y) = 0, \\ &[\gamma = -P; \alpha = -(l + \nu); \beta = -(\lambda + n)]. \end{aligned} \tag{13}$$

(13) demonstrates the first theorem of Jonas: *In our hypothesis the congruence of the lines joining two corresponding points $x$, $y$ is a $W$-congruence; and $u$, $v$ are the asymptotic parameters of its focal surfaces.*

We will now demonstrate the converse of this proposition, i.e. the second theorem of Jonas. Let $z = z(u, v)$ generate the surface $S_z$ and let $u$, $v$ be the asymptotic parameters on this surface. It is known[2] that we can find five functions $\theta$, $\beta$, $\gamma$, $p_{11}$, $p_{22}$ such that

$$z_{uu} = \theta_u z_u + \beta z_v + p_{11} z; \qquad z_{vv} = \gamma z_u + \theta_v z_v + p_{22} z. \tag{14}$$

The most general $W$-congruence, for which $S_z$ is a focal surface, is generated by the lines joining the point $z$ to the point $Az_u + Bz_v$, where $A$, $B$ are two functions different from zero satisfying the equations[3]

$$A_v = -B\gamma; \qquad B_u = -A\beta. \tag{15}$$

---

[2] Cf. Fubini and Čech: *Geometria proiettivo differenziale* (Bologna, Zanichelli) §16, p. 90.

[3] Loc. cit. p., ibid. §§41–42.

If $x$, $y$ are two distinct points on each line $(z, Az_u + Bz_v)$, it is possible to find two functions $\rho$, $\sigma$ such that $\rho \neq \sigma$ and

$$x = \rho z + 2(A_u + A\theta_u), \qquad y = \sigma z + 2(B_v + B\theta_v).$$

Therefore

$$\begin{aligned} x_u &= (\rho_u + 2Ap_{11})z + mz_u + 2Bz_{uv}, \\ x_v &= (\sigma_v + 2Bp_{22})z + nz_v + 2Az_{uv}, \end{aligned} \tag{16}$$

where

$$m = \rho + 2(A_u + A\theta_u); \qquad n = \sigma + 2(B_v + 2B\theta_v). \tag{17}$$

The lines $(x, x_u)$ and $(y, y_v)$ meet if and only if

$$m + n = 0; \text{ i.e. } \rho + \sigma = -2\{(A_u + A\theta_u) + (B_v + B\theta_v)\}. \tag{18}$$

If (18) is satisfied, the lines $(x, x_v)$ and $(y, y_u)$ will intersect. The geometrical significance of this equation is that *the points $x$, $y$ are conjugate with respect to the foci $z$ and $z_1$ of the line $(x, y)$*. It is known[4] that

$$z_1 = -\{(A_u + A\theta_u) + (B_v + B\theta_v)\}z + 2(Az_u + Bz_v).$$

*If $(x, y)$ are conjugate with respect to the foci $z$ and $z_1$ of $(x, y)$, and if the congruence of the lines $(x, y)$ is $W$, then the lines $(x, x_u)$ and $(y, y_v)$ meet, and also the lines $(x, x_v)$ and $(y, y_u)$ intersect.* ($u$ and $v$ are the asymptotic parameters of the focal surfaces.)

Is it possible that the plane $\pi'$ osculating at $x$ the curve $v = \text{const.}$ on $S_x$ is identical with the plane $\pi''$ osculating at $y$ the curve $u = \text{const.}$ on $S_y$? In this case the two planes will coincide with the plane $\pi$ passing through the points $x$, $y$, $x_u$, $y_v$ [which lie on the same plane, because the lines $(x, x_u)$ and $(y, y_v)$ meet]. This plane $\pi$ is obviously the plane through the points $z$, $Az_u + Bz_v$, $x_u$, $y_v$. It will coincide with the planes $\pi'$ and $\pi''$ if and only if the points $x_{uu}$ and $y_{vv}$ lie on $\pi$. By differentiating (16) we find that[5]

$$\begin{aligned} x_{uu} = \Lambda z &+ [\rho_u + m_u + m\theta_u + 2B(\theta_{uv} + \beta\gamma)]z_u \\ &+ [m\beta + 2B(\beta_v + \beta\theta_v)]z_v + 2(B_u + B\theta_u)z_{uv} + 2p_{11}(Az_u + Bz_v), \end{aligned}$$

where it is not necessary to write the value of $\Lambda$.

This point $x_{uu}$ lies on $\pi$ if and only if

$$\rho_u + m_u + 2B(\theta_{uv} + \beta\gamma) - 2A(\beta_v + \beta\theta_v) = 0.$$

This equation, by means of (18) and of (15), becomes:

$$\frac{\partial}{\partial u}(\rho - \sigma) = 0. \tag{19}$$

---

[4] Loc. cit. p., ibid. §41, p. 244 [recall (15)].

[5] Recall equations (15).

In the same way we see that the osculating plane at the point $x$ to the curve $u =$ const. on $S_x$ is identical with the osculating plane at the point $y$ to the curve $v =$ const. on $S_y$, if and only if

$$(19)_{bis} \qquad \frac{\partial}{\partial v}(\rho - \sigma) = 0.$$

From (19) and $(19)_{bis}$ we deduce $\rho - \sigma = 2K =$ const., and from (17) and (18) we get

$$(20) \qquad \begin{aligned} \rho &= K - \{(A_u + A\theta_u) + (B_v + B\theta_v)\}, \\ \sigma &= -K - \{(A_u + A\theta_u) + (B_v + B\theta_v)\}, \end{aligned}$$

or

$$(21) \qquad x = Kz + z_1\,; \qquad y = -Kz + z_1\,, \qquad (K = \text{const.})$$

where $K =$ const. is arbitrary. We deduce the second theorem of Jonas:

*If $u$, $v$ are the asymptotic parameters on the focal surfaces of a $W$-congruence $K$, we can choose on every generator $r$ of $K$ two points $x$, $y$* [conjugate with respect to the foci of $r$] *such that the plane osculating at $x$ the curve $u =$ const.* [$v =$ const.] *of the surface $S_x$ locus of the point $x$, is identical to the plane osculating at $y$ the curve $S_y =$ const.* [$u =$ const.] *of $S_y$.* And the choice of the points $x$, $y$ depends only on *one* arbitrary constant.

We will now give another demonstration of this theorem. It is easy to see that, if $x = x(u, v)$, $y = y(u, v)$, and if $z$, $t$ are the foci on $(x, y)$ of the congruence generated by these lines $(x, y)$, these foci $z$, $t$ are conjugate with respect to $x$, $y$ if and only if

$$(22) \qquad \Gamma = (x_u\,, y_v\,, x, y) + (y_u\,, x_v\,, x, y) = 0.^{6}$$

If the congruence is $W$, and $u$, $v$ are the asymptotic parameters of the focal surfaces, the coördinates $(x, y)$ of its generators satisfy an equation (13), i.e. an equation

$$(x, y_{uv}) + (x_{uv}\,, y) + (x_u\,, y_v) + (x_v\,, y_u) + \alpha[(x_u\,, y) + (x, y_u)] + \beta[(x_v\,, y) + (x, y_v)] + \gamma(x, y) = 0.$$

Multiplying by $(x, y)$ and summing, we get

$$\sum (x, y)(x_u\,, y_v) + \sum (x, y)(x_v\,, y_u) = 0,$$

i.e.

$$(23) \qquad (x, y, x_u\,, y_v) + (x, y, x_v\,, y_u) = 0.$$

If $\Gamma = 0$, from (23) we deduce

$$(24) \qquad (x, y, x_u\,, y_v) = (x, y, x_v\,, y_u) = 0;$$

[6] This equation is identical with the preceding equation (2).

the lines $(x, x_u)$ and $(y, y_v)$ [$(x, x_v)$ and $(y, y_u)$] meet. This is the same theorem we had deduced before from (18).

By a suitable choice of the factors of proportionality we can write

$$x = z + \rho t, \qquad y = z - \rho t, \tag{25}$$

where $\rho$ is a function of $u$, $v$, and (24) becomes equivalent to the equations:

$$\begin{aligned} (z_u + \rho t_u\,,\ z_v - \rho t_v\,,\ z,\ t) &= 0, \\ (z_u - \rho t_u\,,\ z_v + \rho t_v\,,\ z,\ t) &= 0. \end{aligned} \tag{26}$$

From the hypothesis that $z$, $t$ are the foci of the line $(z, t)$ we deduce

$$(z, t, z_u\,, z_v) = (z, t, t_u\,, t_v) = 0, \tag{27}$$

because $z$ lies on the tangent plane of the focal surface $S_t$ and $t$ on the tangent plane of $S_z$. It follows that we can find three functions $\mu$, $A$, $B$ of $u$, $v$, such that

$$t = \mu z + 2(Az_u + Bz_v) \tag{28}$$

and it is obvious that conversely

$$Nz = \bar{\mu} t + 2(At_u - Bt_v) \tag{28}_{\text{bis}}$$

where $N$, $\bar{\mu}$ are other functions of $u$, $v$. The curves $u = \text{const.}$ and $v = \text{const.}$ are asymptotic on both the focal surfaces $S_z$, $S_t$; it follows that we can determine other functions $\delta$, $\theta$, $p_{ii}$, $\beta$, $\gamma$, $b$, $c$, $q_{ii}$ such that

$$\begin{aligned} z_{uu} &= \theta_u z_u + \beta z_v + p_{11} z, & t_{uu} &= \delta_u t_u + b t_v + q_{11} t, \\ z_{vv} &= \gamma z_u + \theta_v z_v + p_{22} z, & t_{vv} &= c t_u + \delta t_v + q_{22} t. \end{aligned} \tag{29}$$

The osculating plane $\pi'$ to the curve $v = \text{const.}$ of the surface $S_x$, locus of the point $x$, is the plane passing through the points

$$x = z + \rho t, \qquad x_u = (z + \rho t)_u\,, \tag{30}$$

$$\begin{aligned} x_{uu} = (z + \rho t)_{uu} &= z_{uu} + 2\rho_u t_u + \rho_{uu} t + \rho t_{uu} \\ &= \theta_u(z_u + \rho t_u) + \beta(z_v - \rho t_v) + p_{11} z \\ &\quad + (\rho q_{11} + \rho_{uu}) t + \rho(b + \beta) t_v + [2\rho_u + (\delta_u - \theta_u)] t_u\,. \end{aligned} \tag{31}$$

The plane $\pi'$ is identical with the plane $\pi''$ osculating the curve $u = \text{const.}$ of $S_y$ if and only if both the planes $\pi'$ and $\pi''$ coincide with the plane $\pi$ of the points

$$x = z + \rho t, \quad y = z - \rho t, \qquad x_u = (z + \rho t)_u\,, \quad y_v = (z - \rho t)_v\,, \tag{32}$$

i.e. with the plane $\pi$ of the points

$$z,\ t,\ z_u + \rho t_u\,,\ z_v - \rho t_v\,. \tag{33}$$

The planes $\pi$, $\pi'$ will be identical if and only if $x_{uu}$ is a linear combination of the points (33); by means of (31) it is easy to prove that $x_{uu}$ is such a combination if and only if the following linear combination of $t$, $t_u$

$$\left(\delta_u - \theta_u + 2\frac{\rho_u}{\rho}\right)t_u + (b+\beta)t_v \tag{34}$$

is equal to such a linear combination [of the points (32)]

$$Lz + Mt + P(z_v - \rho t_v) + Q(z_u + \rho t_u). \tag{35}$$

If we substitute in (35) for $z$, $z_u$, $z_v$ the value $(28)_{bis}$ and the values deduced by differentiation, we see that (35) is equal to an expression

$$at + \bar{b}t_u + \bar{c}t_v + \frac{2}{N}(AP - QB)t_{uv}, \tag{36}$$

where it is not necessary to calculate the values of $a$, $\bar{b}$, $\bar{c}$. If (36) and (34) must be equal to each other, it is necessary that $AP - QB = 0$; and therefore

$$P = 2B\sigma, \qquad Q = 2A\sigma,$$

where $\sigma$ is an arbitrary function. With this value of $P$, $Q$, (35) becomes equal to

$$\left\{\begin{aligned} &Lz + Mt + 2\sigma(Az_u + Bz_v) + 2\sigma\rho(At_u - Bt_v)\\ &\quad = Lz + Mt + \sigma(t - \mu z) + 2\sigma\rho(At_u - Bt_v)\\ &\quad = (L - \sigma\mu)\frac{\bar{\mu}t + 2(At_u - Bt_v)}{N} + (M+\sigma)t + 2\rho\sigma(At_u - Bt_v)\\ &\quad = \left[\frac{\bar{\mu}(L - \sigma\mu)}{N} + M + \sigma\right]t + 2\left(\frac{L - \sigma\mu}{N} + \sigma\rho\right)(At_u - Bt_v), \end{aligned}\right. \tag{37}$$

and if (35) [i.e. the last member of (37)] is equal to (34), it is obvious that

$$\frac{\delta_u - \theta_u + 2\frac{\rho_u}{\rho}}{A} = -\frac{b+\beta}{B},$$

i.e.

$$2\frac{\rho_u}{\rho} = \theta_u - \delta_u - \frac{A}{B}(b+\beta), \tag{38}$$

and this condition is obviously also sufficient. In the same way we obtain: The planes $\pi$, $\pi''$ are identical if and only if

$$2\frac{\rho_v}{\rho} = \theta_v - \delta_v - \frac{B}{A}(c+\gamma). \tag{$38_{bis}$}$$

[This equation is quite analogous to the preceding (38).] From the general theory of congruences $W$ it follows that (38) and $(38)_{bis}$ form an integrable

system,[7] and that therefore $\rho$ exists and depends only on one arbitrary constant. And the second theorem of Jonas is again completely demonstrated.

The surfaces $S_x$ and $S_y$ are characterized by the following analytical property, which is equivalent to the geometrical property, from which we started. The $x$, $x_u$, $x_{uu}$ are linear combinations of the $y$, $y_v$, $y_{vv}$ and conversely the $x$, $x_v$, $x_{vv}$ are linear combinations of the $y$, $y_u$, $y_{uu}$. *A priori* we would not expect that this property is connected with the theory of $W$-congruences.

We state now a new problem which seems to be very important for the theory of $W$-congruences. Let $S_x$ be a given surface. How can we find the system of the lines $u$, $v$ on $S_x$ and another surface $S_y$ in a one-to-one reciprocal correspondence with $S_x$ such that the preceding condition be satisfied?

This problem is a particular case of another: How can we find all the surfaces $S_y$ in a one-to-one correspondence with $S_x$ such that there exist at least two surfaces $S_z$, $S_t$ which are both congruence-transforms of each of the surfaces $S_x$ and $S_y$?

INSTITUTE FOR ADVANCED STUDY

[7] Ibid. Cf. the equations (18) of p. 252. In this equations $b$ and $c$ are indicated by $\bar{\beta}$ and $\bar{\gamma}$.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# ZUR BEGRÜNDUNG DER PROJEKTIVEN GEOMETRIE: EINFÜHRUNG IDEALER ELEMENTE UNABHÄNGIG VON DER ANORDNUNG

VON ARTHUR WINTERNITZ

(Received February 15, 1939)

Die heutige Auffassung, nach welcher jedem beliebigen (nicht notwendig kommutativen) Körper—mag er geordnet sein oder nicht—eine projektive Geometrie entspricht, gibt mir Veranlassung, die zuerst von Klein skizzierte Begründungsweise der projektiven Geometrie, bei welcher man von einem beschränkten Raumgebiet ausgeht, einer neuerlichen Untersuchung zu unterwerfen. Klein's Überlegungen bezogen sich auf die reelle dreidimensionale Geometrie, Pasch's sorgfältige Untersuchungen beruhen auf der Anordnung. Mir kommt es hier darauf an, ausschliesslich Inzidenztatsachen in Betracht zu ziehen.

Hilbert's Axiome der Verknüpfung bringen die einfachsten, vom Standpunkt der Erfahrung aus naheliegenden Annahmen über das Ineinanderliegen von Punkten, Geraden und Ebenen zum Ausdruck. Die projektive Geometrie (des Dreiraumes) hat ausschliesslich Aussagen über solche Inzidenzen zum Gegenstand, stellt aber weitergehende Forderungen an sie. Während Hilbert's Verknüpfungsaxiome bereits in Teilbereichen eines projektiven Raumes z. B. in jedem Teilgebiet eines stetigen projektiven Dreiraumes Geltung haben,—wenn man nur jene Geraden und Ebenen berücksichtigt, welche durch das betrachtete Gebiet hindurchgehen—, ist die wesentliche Annahme der projektiven Geometrie, dass zwei Geraden derselben Ebene stets einen Punkt gemein haben, ebenso wie das Parallelenaxiom—in Hilbert's Fassung IV*—, das zu dieser Annahme führt, in einem beschränkten Gebiet ungültig.

An diese Feststellungen knüpft die Fragestellung an, mit welcher wir uns hier beschäftigen. Es sei ein Raum $\mathfrak{R}$ in dem Sinne bekannt, dass die Klassen der "Punkte," "Geraden" und "Ebenen" gegeben sind und festgesetzt ist, welche Elemente verschiedener Klasse "ineinanderliegen" und welche nicht. Es seien die bekannten Verknüpfungsaxiome Hilbert's erfüllt und überdies die Annahme, dass auf jeder Geraden mindestens drei Punkte liegen. Wir fragen nun: Kann man einen Körper $\mathfrak{K}$ derart angeben, dass $\mathfrak{R}$ einen Teil des projektiven Raumes $\mathfrak{P}$ über $\mathfrak{K}$ isomorph ist; d.h. dass den Punkten, Geraden und Ebenen von $\mathfrak{R}$ Punkte, Geraden und Ebenen von $\mathfrak{P}$ so zugeordnet sind, dass das In- und Auseinanderliegen erhalten bleibt? Mit anderen Worten: Kann man $\mathfrak{R}$ zu einem projektiven Raum ergänzen? Ferner: Wenn es einen solchen Körper gibt, wie weit ist er bestimmt?

In Kap. II. werde ich durch Konstruktion eines Beispiels zeigen, dass es Räume $\mathfrak{R}$ gibt, welche sich nicht zu einem projektiven Raum ergänzen lassen.

So gelangen wir zu der weiteren Frage: Welche zusätzlichen Voraussetzungen sind erforderlich, um die Ergänzbarkeit zu verbürgen? Diese Voraussetzungen sollen dabei gleichfalls Erfahrungen in einem beschränkten Gebiet entsprechen und sich auf Inzidenz allein beziehen. Es ergeben sich zwei notwendige Bedingungen, welche in Kap. III. auch als hinreichend erwiesen werden: H1) "Vierkanthypothese," H2) Desarguesscher Satz im Bündel. Diese Voraussetzungen werden sodann in Kap. IV. auf bedeutend schwächere hinreichende Bedingungen reduziert. Dieselben bringen Sachverhalte zum Ausdruck, welche durch Umgebungsvorstellungen suggeriert werden, aber rein projektiven Charakter haben; z.B. gelten diese meine "Umschliessungs"-Annahmen in den endlichen projektiven Geometrien.

Nichts desto weniger führt mein Ergebnis zu bemerkenswerten topologischen Konsequenzen: Sind in einem Raume $\mathfrak{R}$ einerseits Inzidenzen gemäss Hilbert's Axiomen der Verknüpfung, andererseits Umgebungen, welche die Hausdorffschen Axiome (einschliesslich des ersten Abzählbarkeitsaxioms) erfüllen, erklärt, so genügen naheliegende Stetigkeitsannahmen, um die Ergänzbarkeit von $\mathfrak{R}$ zu einem projektiven Raum zu sichern. Dieser wird dann zu jenen stetigen projektiven Räumen gehören, welche Kolmogoroff und Pontrjagin in Annals of Mathematics (2)33(1932) untersucht haben. Die in Betracht kommenden Körper werden dort unter Benutzung eines Resultats von Frobenius auf eines der drei hyperkomplexen Systeme über dem Körper der reellen Zahlen reduziert: reelle Zahlen, gewöhnliche komplexe Zahlen, Quaternionen.

Bezüglich der Einordnung der auf der Anordnung beruhenden Theorie genügt es zu bemerken, dass meine "Umschliessungs"-Annahmen leicht beweisbare Konsequenzen der Anordnungsaxiome sind.

In dem vorbereitenden Kap. I. habe ich einige Axiomensysteme der projektiven Geometrie und solche, welche Hilbert's Verknüpfungsaxiomen äquivalent sind, einander gegenübergestellt. Besonders übersichtlich wird das Verhältnis durch die Elimination der Geraden, welche bisher meines Wissens in axiomatischen Untersuchungen dieser Art nicht vorgekommen ist.

Einige Bemerkungen zur Geschichte des Gegenstandes findet man im Anhang. Dorthin habe ich auch einige ergänzende Zusätze verwiesen, welche aus dem Text ausgeschieden wurden, um den Fortgang der Entwicklungen nicht zu stören.

## Kapitel I. Axiomensysteme der projektiven Geometrie und Hilbert's Verknüpfungsaxiome

Ausgehend von einem besonders einfachen Axiomensystem der projektiven Geometrie $\mathfrak{a}$, welche sich ausschliesslich auf die Relation "Inzidenz von Punkt und Ebene" bezieht (I.1.), gelangen wir durch Abschwächung zu einem Axiomensystem $\mathfrak{h}$, dessen Aussagen Erfahrungstatsachen entsprechen. (I.2.) Dieses Axiomensystem genügt zur Definition von "Geraden" und ihrer "Inzidenz" mit Punkten und Ebenen. (I.3.) Indem wir nunmehr die Geraden als gleichberechtigte Elemente anerkennen, erhalten wir an Stelle von $\mathfrak{h}$ ein Axio-

mensystem $\mathfrak{H}$, welches leicht als äquivalent mit dem Hilbertschen System der Verknüpfungsaxiome zu erkennen ist. (I.4.) Durch denselben Vorgang gelangen wir zu einem neuen Axiomensystem der projektiven Geometrie $\mathfrak{A}$, das zwar mehr, aber einfachere Axiome enthält als das frühere $\mathfrak{a}$. (I.5.) Schliesslich führen wir auch Veblen's Axiomensystem $\mathfrak{V}$ der projektiven Geometrie an, das sich ausschliesslich auf die Relation "Inzidenz von Punkt und Gerade" bezieht. (I.6.)

In jedem Axiomensystem haben wir unabhängige Annahmen zusammengestellt.

### I.1. **Axiomensystem der projektiven Geometrie, ausgedrückt durch "Inzidenz von Punkt und Ebene"** ($\mathfrak{a}$)

Zur Kennzeichnung der projektiven Geometrie des dreidimensionalen Raumes über einem Körper unabhängig von der Natur des Körpers genügen die folgenden Axiome, welche sich auf eine Relation zwischen den Elementen von zwei Klassen beziehen. Die Elemente der ersten Klasse nennen wir "Punkte" $A, B, C, \cdots$, die der zweiten Klasse "Ebenen" $\alpha, \beta, \gamma, \cdots$. Die Relation, welche zwischen einem Punkt $A$ und einer Ebene $\alpha$ bestehen oder nicht bestehen soll, nennen wir "Inzidenz":

$A\alpha$ "$A$ liegt in $\alpha$" oder "$\alpha$ geht durch $A$ hindurch";

Negation $\overline{A\alpha}$ "$A$ liegt nicht in $\alpha$" oder "$A$ liegt ausserhalb $\alpha$."

Axiom a) (*Inzidenzaxiom*): *Liegen zwei verschiedene Punkte $A$ und $B$ in zwei verschiedenen Ebene $\alpha$ und $\beta$, der Punkt $C$ gleichfalls in $\alpha$ und $\beta$, während die Ebene $\gamma$ durch $A$ und $B$ hindurchgeht, so liegt $C$ in $\gamma$.*

Axiome b), b*) (*Verknüpfungsaxiome*):

Axiom b) *Zu irgend drei Punkten gibt es stets eine Ebene, welche durch sie hindurchgeht.*

Axiom b*) *Zu irgend drei Ebenen gibt es stets einen Punkt, welcher in ihnen liegt.*

Axiom c) (*Fundamentalfigur*) *Man kann fünf Punkte $P_1, P_2, P_3, P_4, P_5$ und fünf Ebenen $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5$ angeben, sodass $P_i$ dann und nur dann ausserhalb $\epsilon_k$ liegt, wenn $i - k \equiv \pm 2 (\mathrm{mod}\ 5)$ ist.*

Das Inzidenzaxiom a) kann auch kurz so ausgesprochen werden: Von den für drei verschiedene Punkte und drei verschiedene Ebenen in Betracht kommenden (neun) Inzidenzen ist eine Konsequenz der übrigen. Es ist gleichwertig mit jeder der beiden Eindeutigkeitsaussagen:

a') *Liegen zwei verschiedene Punkte $A, B$ in der Ebene $\gamma$, der Punkt $C$ ausserhalb $\gamma$, so gibt es nicht mehr als eine Ebene, welche durch $A, B$, und $C$ hindurchgeht.* (*keine zwei $\alpha \neq \beta$*)

a'') *Gehen zwei verschiedene Ebenen $\alpha, \beta$ durch den Punkt $C$, die Ebene $\gamma$ nicht durch $C$, so gibt es nicht mehr als einen Punkt, welcher in $\alpha, \beta$ und $\gamma$ liegt.* (*keine zwei $A \neq B$*)

Zu Axiom c) sei bemerkt, dass je vier Punkte der Fundamentalfigur nicht in einer Ebene liegen und vier Ebenen der Figur nicht durch einen Punkt gehen. Drei Punkte der Fundamentalfigur haben nur eine Ebene, drei Ebenen nur

einen Punkt gemein. Dies folgt mit Hilfe von a). Ein Teil der negativen Bedingungen in c) ist entbehrlich. s. Anhang. Doch haben wir die Axiome so formulieren wollen, dass die Gleichberechtigung beider Klassen unmittelbar hervortritt. (Prinzip der Dualtität)

Jedes Axiom ist von den drei übrigen unábhängig. Für a) sieht man das am einfachsten, indem man zur Fundamentalfigur einen in den fünf Ebenen liegenden Punkt $P$ und eine durch die fünf Punkte und $P$ hindurchgehende Ebene $\epsilon$ hinzunimmt. Für b*) zeigt es etwa der affine Raum; für b) folgt es dualistisch. Ohne c) brauchte es überhaupt keine Elemente zu geben.

Es lässt sich zeigen (vgl. I.5.): Hat man zwei Klassen von Elementen und eine Relation zwischen ihnen, welche unserem Axiomensystem $\mathfrak{a}$ genügt, so lässt sich ein (im Allgemeinen nicht kommutativer) Körper angeben, sodass folgendes gilt: Jedem Punkt $P$ entspricht ein Quadrupel nicht gleichzeitig verschwindender Zahlen $x^i$ des Köpers, welche bis auf einen rechts hinzutretenden Faktor $r \neq 0$ bestimmt sind $(x^i r)$ und jeder Ebene $\epsilon$ ein Quadrupel nicht gleichzeitig verschwindender Zahlen $u_i$ des Körpers, welche bis auf einen links hinzutretenden Faktor $l \neq 0$ bestimmt sind $(lu_i)$; der Inzidenz entspricht das Bestehen der Gleichung

$$\sum u_i x^i = 0, \qquad (i = 1, 2, 3, 4).$$

Der Körper ist bis auf Isomorphie bestimmt. Umgekehrt entspricht jedem Körper eine unseren Axiomen genügende Geometrie, isomorphen Körpern entsprechen isomorphie Räume (dieselbe Geometrie).

## I.2. Abschwächung des Axiomensystems mit Rücksicht auf die Erfahrung ($\mathfrak{h}$ und $\mathfrak{h}'$)

Von den angegebenen Axiomen sind nur a), b) und c) von solcher Natur, dass man auf sie durch die Erfahrung innerhalb eines beschränkten Raumes geführt werden kann, wenn man die Begriffe "Punkt," "Ebene," "Inzidenz" in der üblichen Weise auffasst. Sie treffen z.B. in jedem Gebiet eines stetigen projektiven Raumes zu. Das Axiom b*) hingegen ist nur formal durch das Prinzip der Dualität motiviert. Es gibt jedoch Konsequenzen desselben, die wir von unserem empiristischen Standpunkt aus berücksichtigen müssen. In erster Linie die folgenden beiden:

b*1) *Zu jeder Ebene gibt es einen Punkt, der in ihr liegt.*

b*2) *Wenn zwei Ebenen $\alpha \neq \beta$ einen Punkt $A$ gemein haben, so haben sie auch einen weiteren Punkt $B$ gemein.*

Die Unabhängigkeit der beiden Annahmen ist leicht zu erkennen.

Anstelle von c) wollen wir folgende schwächere Annahme in Betracht ziehen:

c1) *Man kann vier Punkte $P_1$, $P_2$, $P_3$, $P_4$ und vier Ebenen $\epsilon_1$, $\epsilon_2$, $\epsilon_3$, $\epsilon_4$ angeben sodass $P_i$ dann und nur dann ausseryalb $\epsilon_k$ liegt, wenn $i = k$ ist.*

So gelangen wir zu dem Axiomensystem $\mathfrak{h}$:

a), b), b*1), b*2), c1).

Wir werden zeigen, dass $\mathfrak{h}$ dem Hilbertschen System der Verknüpfungsaxiome gleichwertig ist.

Das Axiomensystem wird bereits durch die Ecken, Kanten und Flächen eines Tetraeders erfüllt. Wir werden darum meist eine weitere Annahme hinzunehmen:

e) *Wenn zwei Ebenen zwei Punkte gemein haben, so haben sie noch einen dritten Punkt gemein.*

Das so erweiterte Axiomensystem heisse $\mathfrak{h}'$. Auch e) ist eine Konsequenz von $\mathfrak{a}$. (Beweis in I.5.)

### I.3. Definition der "Geraden" auf Grund von $\mathfrak{h}$, genauer a), b), c1)

Ein Punktepaar heisse mit einem Ebenenpaar "assoziiert," wenn die Elemente des einen Paars mit denen des anderen Paars inzidieren. (Ein Paar bestehe aus zwei verschiedenen Elementen.) Zwei Punktepaare mögen "geradengleich" heissen, wenn es ein Ebenenpaar gibt, das mit beiden assoziiert ist. Diese Beziehung ist symmetrisch und auf Grund von a) transitiv. Die Existenzannahmen b) und c1) ermöglichen es zu jedem Punktepaar $(A, B)$ ein assoziiertes Ebenenpaar zu finden. Ein solches reicht zur Entscheidung hin, ob ein Punktepaar zu $(A, B)$ geradengleich ist. (Axiom a)) Die Klasse aller zum Punktepaar $(A, B)$ geradengleicher Punktepaare werde "Gerade $AB$" genannt, jede solche Klasse eine "Gerade." Ein Punkt $P$ liegt auf der Geraden $g$, wenn er einem Paar der zu $g$ gehörigen Klasse angehört. Es ergibt sich leicht: Drei Punkte liegen dann und nur dann auf einer Geraden, wenn sie in zwei verschiedenen Ebenen enthalten sind. Die "Inzidenz" von Gerade und Ebene kann so erklärt werden: Ist $(A, B)$ ein Punktepaar der Klasse $g$, so "liegt" $g$ in $\epsilon$, wenn $A$ und $B$ beide in $\epsilon$ liegen. Auf die Auswahl kommt es nicht an.

### I.4. Hilbert's Verknüpfungsaxiome, durch Inzidenz von Punkten, Geraden und Ebenen ausgedrückt ($\mathfrak{H}$)

Indem wir die "Gerade" unter die Grundelemente aufnehmen und ihre "Inzidenz" mit Punkt und Ebene als Grundbeziehung auffassen, erhalten wir an Stelle von $\mathfrak{h}$ das Axiomensystem $\mathfrak{H}$:

Axiom A0) *Inzidiert der Punkt $A$ und die Ebene $\alpha$ mit derselben Geraden $a$, so inzidieren $A$ und $\alpha$ miteinander.*

Axiom A1) *Liegt der Punkt $A$ ausserhalb der Geraden $a$, so gibt es höchstens eine Ebene, welche mit beiden inzidiert.*

Axiom A2) *Zu zwei verschiedenen Punkten $A$, $B$ gibt es höchstens eine Gerade, welche mit beiden inzidiert.*

Axiom B1) *Ein Punkt $A$ und eine Gerade $a$ ($A$ ausserhalb $a$) haben stets eine Ebene gemein.*

Axiom B1*1) = b*1) *Zu jeder Ebene gibt es einen Punkt, der in ihr liegt.*

Axiom B1*2) *Zu jeder Geraden gibt es zwei verschiedene Punkte, die auf ihr liegen.*

Axiom B2*1) *Wenn zwei Ebenen $\alpha \neq \beta$ einen Punkt $A$ gemein haben, so haben sie eine Gerade gemein.*

Axiom C1) = c1) *Man kann vier Punkte* $P_1$, $P_2$, $P_3$, $P_4$ *und vier Ebenen* $\epsilon_1$, $\epsilon_2$, $\epsilon_3$, $\epsilon_4$ *angeben, sodass* $P_i$ *dann und nur dann ausserhalb* $\epsilon_k$ *liegt, wenn* $i = k$ *ist.*

Aus diesen Axiomen, welche unabhängig sind, ergeben sich die Folgerungen:

A1*) *Liegt die Gerade a nicht in der Ebene* $\alpha$, *so haben sie höchstens einen Punkt gemein.*

A2*) *Zwei verschiedene Ebenen* $\alpha$, $\beta$ *haben höchstens eine Gerade gemein.*

B2) *Zwei Punkte* $(A \neq B)$ *haben stets eine Gerade gemein.*

Bezüglich der Beweise siehe Anhang; ebenso bezüglich des Beweises der Unabhängigkeit der Axiome des Systems $\mathfrak{H}$.

Nun ergibt sich ohne weiteres die Gleichwertigkeit von $\mathfrak{h}$ bezw. $\mathfrak{H}$ mit Hilbert's Axiomgruppe I (*Grundlagen der Geometrie*, Kap. I, 7. Aufl., Leipzig 1930). Dort ist die Zusammengehörigkeit von Gerade und Ebene nicht unter die Grundbegriffe aufgenommen. S. Anhang.

Schliesslich nimmt noch unsere Zusatzannahme e) die Gestalt an:

E) *Zu jeder Geraden gibt es wenigstens drei verschiedene Punkte, die auf ihr liegen.*

Das Axiomensystem $\mathfrak{h}'$ geht über in $\mathfrak{H}'$.

### I.5. Axiomensystem der projektiven Geometrie, durch Inzidenz von Punkten, Geraden und Ebenen ausgedrückt ($\mathfrak{A}$)

Durch die Aufnahme der "Geraden" unter die Grundelemente und ihrer "Inzidenzen" unter die Grundbeziehungen nimmt unser Axiomensystem der projektiven Geometrie $\mathfrak{a}$ eine Gestalt $\mathfrak{A}$ an, in der nur einfache Aussagen über Verknüpfungen von zwei Elementen vorkommen.

A0), A1), A2), B1) werden beibehalten; dazu kommen:

Axiom B1*) *Eine Gerade a und eine Ebene* $\alpha$ *(welche nicht ineinanderliegen) haben stets einen Punkt gemein.*

Axiom B2*) *Zwei (verschiedene) Ebenen* $\alpha$, $\beta$ *haben stets eine Gerade gemein.*

Axiom C) = c) (s.I.1.)

Aus $\mathfrak{A}$ folgt offenbar $\mathfrak{H}$, und somit die dualen Ergänzungen: A1*), A2*), B2).

Ferner folgt E) und der duale Satz:

E*) *Zu jeder Geraden gibt es drei Ebenen, die mit ihr inzidieren.*

Zum Beweis des letzten Satzes vervollständige man die Fundamentalfigur c), indem man je drei ihrer fünf Ebenen zum Schnitt bringt. Die erhaltenen zehn Punkte lassen sich dann nicht in zwei Ebenen $\alpha$, $\beta$ unterbringen. Von den fünf Punkten $P$ aus c) müssten drei in eine der Ebenen $(\alpha)$ fallen, welche also zu den $\epsilon$ aus c) gehört. Die übrigen vier Ebenen $\epsilon$ schneiden sich ausserhalb $\alpha$ in den Ecken eines Tetraeders. Diese können nicht in $\beta$ untergebracht werden. Unter den Verbindungsebenen von $g$ mit diesen zehn Punkten gibt es also sicher drei verschiedene.

Hervorgehoben seien noch die Folgerungen:

G), G*) *Wenn zwei Geraden einen Punkt gemein haben, so haben sie auch eine Ebene gemein und umgekehrt.*

$\mathfrak{A}$ besteht aus unabhängigen Axiomen. Beweis im Anhang.

### I.5.1. Zurückführung auf einen Körper

Nun sind alle Voraussetzungen gegeben, um im Wesentlichen nach Hilbert (*Grundl. d. Geom.* Kap. V.) die Zurückführbarkeit der projektiven Geometrie auf einen Körper zu zeigen. Man kann so vorgehen, dass man eine Ebene und die mit ihr inzidierenden Elemente ausschliesst. Dann bleibt ein "affiner" Raum zurück, d.h. ein Raum, in welchem Hilbert's Axiome der Verknüpfung (I.) und das Parallelenaxiom (IV*) gelten. Nun lässt sich, obwohl die Anordnungsaxiome fehlen, der Vektor-Begriff nach dem Vorgange von O. Hölder in "Streckenrechnung und projektive Geometrie," Leipz. Ber. 1911, begründen und das Rechnen mit Verhältnissen paralleler Vektoren einführen, sowie die Rechengesetze bis auf das Kommutativgesetz der Multiplikation beweisen. (Hilbert, *Grundl. d. Geom.* Kap. III, Gesetze 1–11) Nun ergibt sich auch die am Schluss von I.1. behauptete Linearität der Ebenengleichung.

### I.6. "Inzidenz von Punkt und Gerade" als Grundbegriff der projektiven Geometrie. (Axiomensystem $\mathfrak{B}$)

Aus dem Axiomensystem der projektiven Geometrie $\mathfrak{A}$ kann man auch den Begriff der Ebene eliminieren und zu einem System gelangen, das sich ausschliesslich auf die "Inzidenz von Punkt und Gerade" bezieht, wie es etwa Veblen und Young in ihrem Werk "Projective Geometry" (Ginn and Comp., 1910, repr. 1916) zum Ausgangspunkt ihrer Betrachtungen machen. Das dort in Kap. I. angegebene Axiomensystem lautet:

DEFINITION DER EBENE: *Sind $P$, $Q$, $R$ drei nicht auf derselben Geraden gelegene Punkte und ist $g$ eine Gerade, welche $Q$ und $R$ verbindet; so wird die Klasse $\mathfrak{P}_2$ aller Punkte auf den Geraden, welche $P$ mit den Punkten von $g$ verbinden "die durch $P$ und $g$ bestimmte Ebene" genannt.*

DEFINITION DES DREIRAUMS: *Sind $P$, $Q$, $R$, $S$ vier nicht in derselben Ebene gelegene Punkte und ist $\epsilon$ eine Ebene durch $Q$, $R$ und $S$; so wird die Klasse $\mathfrak{P}_3$ aller Punkte auf den Geraden, welche $P$ mit den Punkten von $\epsilon$ verbinden, "der durch $P$ und $\epsilon$ bestimmte Dreiraum" genannt.*

DIE ANNAHMEN DER INZIDENZ (ALIGNMENT):

A1. *Wenn $A$ und $B$ verschiedene Punkte sind, so gibt es wenigstens eine Gerade durch $A$ und $B$ zugleich.*

A2. *Wenn $A$ und $B$ verschiedene Punkte sind, so gibt es nicht mehr als eine Gerade durch $A$ und $B$ zugleich.*

A3. *Sind $A$, $B$, $C$ nicht auf derselben Geraden gelegene Punkte und $D \neq E$ solche Punkte, dass $B$, $C$, $D$ auf einer Geraden und $C$, $A$, $E$ auf einer Geraden liegen; so gibt es einen solchen Punkt $F$, dass $A$, $B$, $F$ auf einer Geraden und auch $D$, $E$, $F$ auf einer Geraden liegen.*

DIE ANNAHMEN DER AUSDEHNUNG (EXTENSION):

E0. *Es gibt wenigstens drei Punkte auf jeder Geraden.*

E1. *Es gibt wenigstens eine Gerade.*

E2. *Nicht alle Punkte liegen auf derselben Geraden.*

E3. *Nicht alle Punkte liegen in derselben Ebene.*

ANNAHME DER GESCHLOSSENHEIT (CLOSURE):

$\overline{E4}$. *Alle Punkte liegen in demselben Dreiraum.*

(Nur das letzte Axiom habe ich ein wenig abgeändert.)

Die Gleichwertigkeit dieses Axiomensystems $\mathfrak{B}$ mit dem früheren System $\mathfrak{A}$ geht aus Kap. I. des zitierten Werkes hervor. Das Axiomensystem $\mathfrak{B}$ entfernt sich in seiner einschneidenden Annahme A3. besonders weit von der empirischen Plausibilität. Wir kommen auf dasselbe nach Einführung "idealer" Punkte und Geraden zurück.

## KAPITEL II. FÜHREN HILBERT'S VERKNÜPFUNGSAXIOME ALLEIN AUF EINEN KÖRPER?

### II.1. Eine notwendige Bedingung für die Ergänzbarkeit eines den Hilbertschen Verknüpfungsaxiomen genügenden Raumes zu einem projektiven: Vierkanthypothese H1)

Unter einem "Dreikant" $\Delta$ verstehen wir die Figur von drei Ebenen $\alpha, \beta, \gamma$, welche keine Gerade (höchstens einen Punkt) gemein haben, aber sich paarweise schneiden. Die drei Schnittgeraden $a, b, c$ ("Kanten") liegen nicht in einer Ebene, aber je zwei sind in einer Ebene enthalten. ($b$ and $c$ in $\alpha$ usw.) Der etwaige gemeinsame Punkt von $\alpha, \beta$ und $\gamma$ heisst "Scheitel." (Durch ihn gehen $a, b, c$ hindurch.)

Eine notwendige Bedingung für die Ergänzbarkeit eines Raumes $\mathfrak{R}$, in welchem das Axiomensystem $\mathfrak{H}$ gilt, zu einem projektiven Raum ist offenbar die folgende:

H1) (*Vierkanthypothese*) *Haben zwei Dreikante zwei Kanten gemein* ($a, b, c$ *und* $a, b, d$), *so sind die dritten Kanten* ($c$ *und* $d$) *in einer Ebene gelegen* ("*komplanar*").

In der Tat müssen im erweiterten Raume die Scheitel der beiden Dreikante mit dem Schnittpunkt von $a$ und $b$ zusammenfallen.

### II.2. Die Unabhängigkeit der aufgestellten Bedingung. $\mathfrak{h}'$ führt nicht auf einen Körper

Wir werden nun ein Beispiel konstruieren, welches zeigt, dass H1) von $\mathfrak{h}'$ unabhängig ist.

Ausgehend von dem Körper $\mathfrak{K}$ der rationalen Zahlen bilden wir ein komplexes Zahlensystem, das aus Zahlenpaaren $x = (\xi_1, \xi_2)$, $y = (\eta_1, \eta_2)$, $a = (\alpha_1, \alpha_2) \cdots$ besteht. Die Addition erfolgt komponentenweise: $x + y = (\xi_1 + \eta_1, \xi_2 + \eta_2)$, während die Multiplikation $y = ax$ nach einer Regel von der Gestalt erfolgt:

$$\eta_1 = \alpha_1\xi_1 - f(\alpha_2)\xi_2 ,$$
$$\eta_2 = \alpha_2\xi_1 + \alpha_1\xi_2 .$$

Dabei ist die Funktion $f(\alpha)$ folgendermassen in $\mathfrak{K}$ erklärt:

$$\alpha \leqq 0 \rightarrow f(\alpha) = \alpha,$$
$$\alpha \geqq 0 \rightarrow f(\alpha) = 2\alpha.$$

Die Zahlenpaare, deren zweite Komponente verschwindet, bilden ein System, das zu $\mathfrak{R}$ isomorph ist; $(\xi, 0)$ entspricht $\xi$; $(0, 0)$ spielt die Rolle der 0, $(1, 0)$ die der 1. Es gilt das hintere Distributivgesetz: $a(x + y) = ax + ay$, aber nicht das vordere. $(\alpha\beta < 0 \rightarrow f(\alpha + \beta) \neq f(\alpha) + f(\beta))$ Auch das Assoziativgesetz ist ungültig. $y = ax$ ist für $x \neq 0$ nach $a$ auflösbar, weil $f(\alpha) + \kappa^2\alpha$ für jedes $\kappa$ aus $\mathfrak{R}$ alle Werte durchläuft.

Nach dieser Vorbereitung konstruieren wir einen Raum, dessen Punkte durch Zahlentripel $(\xi, x_1, x_2)$ gegeben sind, wobei $\xi$ eine Zahl des Körpers, $x_1$ und $x_2$ komplexe Zahlen im eben erklärten Sinne sind. Die "Ebenen" seien durch lineare Gleichungen von einer der Formen gegeben:

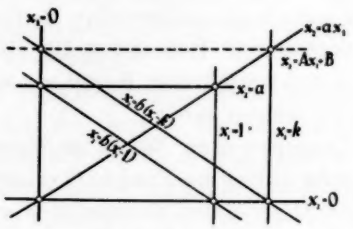
$$x_2 = ax_1 + b\xi + c$$

oder

$$x_1 = b\xi + c$$

oder

$$\xi = \gamma.$$




$(a + b = 0)$

FIG. 1

(Wir haben also teils dreidimensionale, teils vierdimensionale Flächen im fünfdimensionalen Raum.) Das Bestehen der Verknüpfungsaxiome $\mathfrak{h}'$ ergibt sich leicht mit Hilfe der Rechenregeln:

$$(a + b)\xi = a\xi + b\xi,$$

$$ab \cdot \xi = a \cdot b\xi.$$

Um zu zeigen, dass die Vierkanthypothese nicht zutrifft, betrachten wir zunächst die (Nicht-Desarguessche) Geometrie, welche in einer Ebene $\xi = \gamma$ gilt. Die "Geraden" sind durch Gleichungen von der Gestalt

$$x_2 = ax_1 + b \qquad \text{oder} \qquad x_1 = c$$

gegeben. Es gelten die ebenen Verknüpfungsaxiome (Hilbert I., 1–3.) und das Parallelenaxiom (Hilbert IV*). Wir ziehen jetzt die Abhängigkeit der Seitenrichtungen eines vollständigen Vierecks in Betracht. Die Ecken seien: $(0, 0)$, $(k, 0)$, $(0, -bk)$, $(k, ak)$; die Seiten: $x_2 = 0$, $x_1 = 0$, $x_1 = k$, $x_2 = ax_1$, $x_2 = b(x_1 - k)$ und $x_2 = Ax_1 + B$, wobei $-bk = B$, $ak = Ak + B$, also $Ak = ak + bk$ ist. Für $k = 1$ ergibt sich $A = a + b$; für $\kappa_2 \neq 0$ hingegen ergibt sich $A \neq a + b$, wenn $\alpha_2$, $\beta_2$ verschiedene Vorzeichen haben, z.B. $a = (0, 1)$, $b = -a$.

Nun ist eine Gerade in der Ebene $\xi = 0$ zu einer Geraden in der Ebene $\xi = 1$ dann und nur dann komplanar, wenn die Richtungskonstanten übereinstimmen. Betrachten wir das Viereck mit $k = 1$ in der Ebene $\xi = 0$ und das Viereck mit $k = (0, 1)$ in der Ebene $\xi = 1$, so erkennen wir an den nicht übereinstimmenden Richtungskonstanten der letzten Seiten die Ungültigkeit der Vierkanthypothese. (Das erste Dreikant ist durch die Ebenen $x_2 = 0$, $x_1 = 0$, $x_2 = bx_1 + (b - bk)\xi - b$ gegeben; das zweite Dreikant durch die Ebenen $x_2 = 0$, $x_1 = (k - 1)\xi + 1$, $x_2 = ax_1$. Dabei sei $k = (0, 1) = a$, $a + b = 0$.

Hilbert's Verknüpfungsaxiome und die Annahme, dass auf jeder Geraden mindestens drei Punkte liegen, reichen also für die Zurückführbarkeit auf einen Körper nicht hin.

### II.3. Eine zweite notwendige Bedingung für die Ergänzbarkeit eines den Hilbertschen Verknüpfungsaxiomen genügenden Raumes zu einem projektiven: H2) (Desargues)

Da mir nicht bekannt ist, ob die aufgestellte Bedingung hinreicht, gebe ich hier noch eine zweite notwendige Bedingung für die Ergänzbarkeit eines $\mathfrak{h}'$ genügenden Raumes zu einem projektiven Raum an, welche sich zusammen mit der ersten als hinreichend erweisen wird:

H2) *Die Gültigkeit des Desarguesschen Satzes im Bündel.*

Die Desarguessche Konfiguration im Punkte $O$ erhält man ausgehend von der Figur von fünf nicht durch $O$ gehenden Ebenen $[i]$ $(i = 1, 2, 3, 4, 5)$, von denen keine vier durch einen Punkt gehen, indem man—nach Ergänzung zu einem projektiven Raum—die Schnittgeraden $[ij]$ $(i \neq j)$ von je zwei und die Schnittpunkte $[ijk]$ $(i, j, k$ von einander verschieden) von je drei von $O$ aus projiziert. $(i, j, k$ durchlaufen die Werte 1, 2, 3, 4, 5). Der Desarguessche Satz behauptet nun, dass eine der Inzidenzen Konsequenz der übrigen Aussagen ist.

Unser obiges Beispiel (II.2.) zeigt auch die Unabhängigkeit von H2) vom Axiomensystem $\mathfrak{h}'$. In der Tat brauchen wir bloss die beiden Vierecke in die Ebene $\xi = 1$ zu verlegen und vom Nullpunkt aus zu projizieren.

## Kapitel III. Einführung idealer Elemente auf Grund der Hypothesen H1) und H2)

Wir gehen nun darauf aus zu zeigen, dass die beiden Hypothesen H1) und H2) zusammen hinreichend sind, um einen dem Axiomensystem $\mathfrak{h}'$ oder $\mathfrak{H}'$ genügenden Raum zu einem projektiven Raum ergänzen zu können. Dies geschieht durch Einführung "idealer Elemente."

### III.1. Einführung idealer Punkte auf Grund von H1)

Wir setzen jetzt die Axiome $\mathfrak{h}$ bezw. $\mathfrak{H}$ und H1) (vgl. II.1.) voraus.

Wir erinnern zunächst an die in II.1. gegebene Erklärung eines "Dreikants". Hat man nun ein solches Dreikant $\Delta$ und einen vom etwaigen Scheitel verschiedenen Punkt $P$, so kann man eine Gerade $g$ durch $P$ legen, welche mit jeder der drei Kanten $a$, $b$, $c$ komplanar ist. Wir sagen in diesem Fall: "$g$ schliesst sich dem Dreikant $\Delta$ an."

Liegt $P$ auf einer Kante $a$, so ist $g = a$; liegt $P$ ausserhalb $a$ und $b$ in deren Verbindungsebene $\gamma$, so verbinde man, $P$ mit $c$ durch eine Ebene, ihr Schnitt mit $\gamma$ ist $g$; liegt $P$ ausserhalb $\gamma$, so verbinde man $P$ mit $a$ bezw. $b$, die beiden Ebenen schneiden sich in einer Geraden $g$, welche auf Grund von H1) auch mit $c$ komplanar ist.

Von einer Ebene $\epsilon$ sagen wir, "sie schliesst sich dem Dreikant $\Delta$ an," wenn $\epsilon$ eine Gerade $g$ enthält, die sich $\Delta$ anschliesst. Ob dies der Fall ist, erkennt man, indem man durch einen beliebigen Punkt von $\epsilon$ die sich $\Delta$ anschliessende Gerade legt.

Ist $ABC$ ein Dreieck, dessen Ebene sich dem Dreikant $\Delta$ nicht anschliesst, so kann man ein Dreikant $\Delta'$ angeben, dessen Kanten bezw. durch $A$, $B$, $C$ hindurchgehen und sich dem Dreikant $\Delta$ anschliessen. Schliessen sich alle drei Kanten eines Dreikants $\Delta'$ dem Dreikant $\Delta$ an, so sagen wir: $\Delta'$ ist "scheitelgleich" zu $\Delta$. Dies kommt offenbar darauf hinaus, dass die Kanten des einen Dreikants mit denen des anderen komplanar sind. (9 Relationen)

Es gilt der Satz: *Schliessen sich zwei Geraden einem Dreikant an, so sind sie komplanar.* (Dies folgt aus H1).)

Nun erweist sich die Scheitelgleichheit als symmetrische und transitive Beziehung. Die Dreikante zerfallen in Klassen, welche wir "ideale Punkte" nennen wollen. Die Klasse aller zu $\Delta$ scheitelgleichen Dreikante etwa ist der "ideale Punkt" D. Ist ein Scheitel vorhanden, so besteht die Klasse aus allen Dreikanten mit demselben Scheitel. Auf Grund von Axiom C1) erkennt man ohne weiteres, dass jeder Punkt Scheitel mindestens eines Dreikants ist. Die Einführung der idealen Punkte kann also als Erweiterung des gegebenen Punktraumes aufgefasst werden. Die bisherigen Punkte nennen wir zur Unterscheidung "reale Punkte" und machen sie im Druck besonders kenntlich. (z.B. **D**) Diese Festsetzung gilt für Kap. III. und den Anhang zu IV.2. Hat ein Dreikant den Scheitel **S**, so sagen wir auch: der betreffende ideale Punkt "deckt sich" mit **S**.

Wir erklären nun die Inzidenz einer Geraden $g$ mit einem idealen Punkt $D$ durch die Bedingung: $g$ schliesst sich einem $D$ repräsentierenden Dreikant an; auf die Auswahl desselben kommt es nicht an (H1)). Ein realer Punkt **P**, der sich nicht mit $S$ deckt, gibt also zu einer eindeutig bestimmten Verbindungsgeraden mit $S$ Anlass. Wir sagen ferner: "eine Ebene $\epsilon$ geht durch den idealen Punkt $S$ hindurch" oder "$S$ liegt in $\epsilon$," wenn $\epsilon$ sich einem $S$ repräsentierenden Dreikant anschliesst. Auf die Auswahl kommt es nicht an. Ist $g$ eine nicht durch den idealen Punkt $S$ gehende Gerade, so gibt es also eine und nur eine Ebene, welche mit beiden inzidiert.

Der ideale Punkt $S$ ist offenbar bereits durch zwei Kanten $a$, $b$ oder durch eine Kante $c$ und die gegenüberliegende Seitenfläche $\gamma$ bestimmt. Er kann also durch zwei komplanare Geraden oder durch eine Ebene und eine nicht in ihr liegende Gerade angegeben werden.

Durch drei Punkte **O**, $A$, $B$ von denen wenigstens einer real ist, geht eine Ebene hindurch. (Wenn sie nicht auf einer Geraden liegen, eine einzige; sie enthält die Geraden **O***A* und **O***B*.)

### III.2. Einführung "idealer Geraden" unter Hinzunahme von H2)

Nun tritt uns die entscheidende Frage entgegen, ob das Inzidenzaxiom a) bestehen bleibt, wenn die drei Punkte $A$, $B$, $C$ ideale Punkte sind. Wir wollen dabei wieder die Annahme machen, dass auf jeder Geraden drei reale Punkte liegen ($E$)). Wir setzen also $\mathfrak{h}'$ und H1) voraus.

Wir werden zeigen, dass die gestellte Frage darauf hinauskommt, ob der Desarguessche Satz im Bündel gilt (Hypothese H2), vgl. II. 3).

Es seien $\alpha \neq \beta$ zwei Ebenen, welche mit drei von einander verschiedenen idealen Punkten $A$, $B$, $C$ inzidieren. Wir dürfen annehmen, dass $\alpha$ und $\beta$ keinen realen Punkt gemein haben. Die dritte Ebene $\gamma$ gehe durch $A$ und $B$ hindurch. Ist **O** ein realer Punkt in $\gamma$ ausserhalb $\alpha$ und $\beta$, so kommt es auf die Frage an, ob die drei Geraden **O**$A$, **O**$B$, **O**$C$ in einer Ebene liegen. Dies ist die Bedingung dafür, dass $\gamma$ durch $C$ hindurchgeht. Wir wollen die Ebene $\alpha$ mit [4], die Ebene $\beta$ mit [5] bezeichnen, den Punkt $A$ mit [145], $B$ mit [245], $C$ mit [345] bezeichnen. **O** liegt ausserhalb [4] und [5].

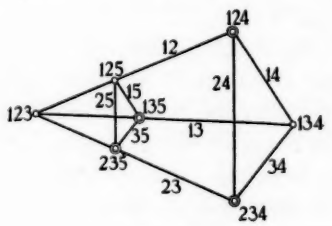



Fig. 2

In [5] wähle den realen Punkt **[235]**. Die Gerade [25] verbinde diesen Punkt mit [245], sie liegt in [5].

In [4] wähle den realen Punkt **[124]**, jedoch ausserhalb der Verbindungsebene **O**[25]. (Möglich auf Grund von $\mathfrak{H}$.) **O** liegt dann ausserhalb der Ebene [2], welche [25] mit **[124]** verbindet.

Die Gerade [14] verbinde **[124]** mit [145] (in [4]);

die Gerade [35] ($\neq$ [25]) verbinde **[235]** mit [345] (in [5]).

Wähle den realen Punkt **[135]** $\neq$ **[235]** auf [35], jedoch ausserhalb der Ebene **O**[14]. (Annahme E)). Die Ebene [1], welche [14] mit **[135]** verbindet, geht also nicht durch **O**.

Die Gerade [24] ($\neq$ [14]) verbinde **[124]** mit [245] (in [4]).

Auf [24] wähle den realen Punkt **[234]** $\neq$ **[124]**, jedoch ausserhalb der Ebene **O**[35]. **O** liegt dann ausserhalb der Ebene [3], welche **[234]** mit [35] verbindet.

Von den fünf Ebenen [1], [2], [3], [4], [5] gehen keine vier durch einen Punkt; **O** liegt ausserhalb von allen.

[2] und [3] schneiden sich in der Geraden [23], [3] und [1] in [31], [1] und [2] in [12]. [1], [2], [3] haben den Punkt [123] gemein. [2], [5], [1] oder [25] und [1] schneiden sich in [125]; [3], [4], [1] oder [34] und [1] in [134]. [15] verbindet **[135]** mit [145], [14] verbindet **[124]** mit [134].

Im Ganzen haben wir eine Desarguessche Konfiguration erhalten: zwei Dreiecke, welche in Bezug auf einen Punkt perspektiv sind. Durch Projektion erhalten wir in **O** zwei Dreikante, welche in Bezug auf einen Strahl perspektiv sind. Der Desarguessche Satz im Bündel gibt unsere Behauptung.

Nun sind wir in der Lage, "ideale Geraden" nach dem Muster von I.3. zu definieren. Ein Ebenenpaar heisse mit einem Paar idealer Punkte assoziiert, wenn Inzidenz zwischen den Elementen des einen und denen des anderen Paars stattfindet. (Ein Paar bestehe aus zwei verschiedenen Elementen.)

Sind $A \neq B$ zwei ideale Punkte, **P**, **Q** zwei reale Punkte; so gibt es eine Ebene $\alpha$ durch **P**, $A$, $B$ und eine Ebene $\beta$ durch **Q**, $A$, $B$. Ist Ebene **PQ**$A \neq$ Ebene **PQ**$B$, so folgt $\alpha \neq \beta$.

Zwei Ebenenpaare heissen "geradengleich," wenn es ein Paar idealer Punkte gibt, mit welchem beide Ebenenpaare assoziiert sind. Diese symmetrische Beziehung erkennen wir jetzt auf Grund der erweiterten Gültigkeit von a) als transitiv. Die Klasse aller zu einem Ebenenpaar $\alpha \neq \beta$ geradengleichen Ebenenpaare nennen wir die "ideale Gerade" $\alpha\beta$. Eine Ebene "inzidiert" mit einer idealen Geraden, wenn sie einem Paar der Klasse angehört. Ein idealer Punkt "inzidiert" mit einer idealen Geraden, wenn er den beiden Ebenen eines ihrer Repräsentanten angehört. Auf die Auswahl kommt es nicht an. Liegt ein realer Punkt in den beiden Ebenen $\alpha$, $\beta$, so bekommt man eine reale Schnittgerade und die Klasse besteht aus allen Ebenenpaaren, welche sich in dieser Geraden schneiden. Wir sagen in diesem Falle: die reale Gerade "deckt sich" mit der idealen.

Es gelten die Sätze: *Inzidiert eine ideale Gerade mit einem Punkt und einer Ebene, so liegt der Punkt in der Ebene. Eine ideale Gerade und eine Ebene haben genau einen Punkt gemein, wenn sie nicht ineinanderliegen.*

*Deckt sich ein idealer Punkt einer idealen Geraden mit einem realen Punkt, so deckt sich die ideale Gerade mit einer realen Geraden.* Kürzer ausgedrückt: *eine Gerade durch einen realen Punkt ist real. Liegt ein realer Punkt auf einer realen Geraden, so liegen auch die idealen Elemente ineinander, welche sich mit ihnen decken.*

### III.3. Gültigkeit des projektiven Axiomensystems $\mathfrak{V}$. ("Ideale Ebenen")

Nun kommt es nur noch darauf an zu zeigen, dass die idealen Punkte und Geraden mit der eben erklärten "Inzidenz von Punkt und Gerade" dem Veblenschen Axiomensystem $\mathfrak{V}$ der projektiven Geometrie (I.6) genügen.

A1. und A2. ergeben sich sofort aus der obigen Definition. Die grundlegende Annahme A3. erfordert eine eingehendere Betrachtung. Sie führt uns auf zwei perspektive Tetraeder. Die Paare entsprechender Ebenen geben die Geraden, die Paare entsprechender Kanten die Punkte der Figur an.

Wir bringen A3. in die Form:

*Zwei Geraden $I \neq II$ mögen einen Punkt $(I\ II)$ gemein haben, durch welchen die Geraden III und IV nicht hindurchgehen, während sie sowohl I als auch II*

*treffen—in den Punkten (I III), (II III) bezw. (I IV), (II IV). Dann haben III und IV einen Punkt (III IV) gemein.*

Die fünf gegebenen Punkte dürfen offenbar als von einander verschieden vorausgesetzt werden. Ferner dürfen wir annehmen, dass es keine (reale) Ebene gibt, in welcher I und II beide liegen.

Nun wähle einen realen Punkt **3** und verbinde ihn mit I, II, IV durch die Ebenen [1], [2], [4]. Sie bilden ein Dreikant.

Hierauf wähle **4*** ausserhalb [1] und ausserhalb [2] und verbinde diesen realen Punkt mit I, II, III durch die Ebenen [1*], [2*], [3*]. Auch sie bilden ein Dreikant.

Ferner wähle den realen Punkt **1** $\neq$ **3** in [2] und [4] zugleich, aber ausserhalb [3*];

überdies den realen Punkt **2*** $\neq$ **4*** in [1*] und [3*] zugleich, aber ausserhalb [4]. (Annahme E))

Es gilt [1*] $\neq$ [1], [2*] $\neq$ [2] wegen ihres verschiedenen Verhaltens zu **4***, [3*] $\neq$ [3] im Hinblick auf **1**, [4*] $\neq$ [4] im Hinblick auf **2***.

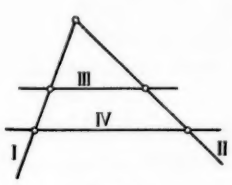



FIG. 3

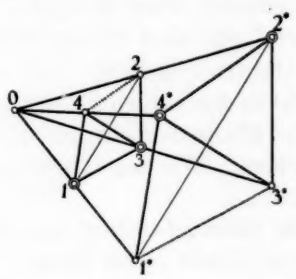



FIG. 4

Sodann verbinde **1** mit III durch [3]. Diese Ebene geht nicht durch **3**, sonst lägen I und II in ihr.

Ferner verbinde **2*** mit IV durch [4*]. Diese Ebene geht nicht durch **4*** hindurch.

Der Schnittpunkt von [3], [1], [4] heisse **2**($\neq$ **2*** wegen [4]),

der Schnittpunkt von [3], [1], [2] heisse **4** ($\neq$ **4*** wegen [2]),

der Schnittpunkt von [4*], [2*], [3*] heisse **1***($\neq$ **1** wegen [3*]),

der Schnittpunkt von [4*], [1*], [2*] heisse **3***($\neq$ **3** weil **34** $\neq$ **3*4***).

Wir haben zwei Tetraeder erhalten: in dem einen liegen den Ebenen [1], [2], [3], [4] bezw. die Ecken **1**, **2**, **3**, **4** gegenüber, in dem anderen liegen den Ebenen [1*], [2*], [3*], [4*] die Ecken **1***, **2***, **3***, **4*** gegenüber. Die Punkte **3**, **4**, **3***, **4*** liegen in einer (realen) Ebene; die komplanaren realen Geraden **33*** $\neq$ **44*** definieren den Punkt *O*. Die Ebenen **133*** und **144*** haben eine reale Schnittgerade **11***, ebenso die (realen) Ebenen **2*33*** und **2*44*** eine reale Schnittgerade **22***. **11*** und **22*** gehen durch *O*; **1**, **2**, **1***, **2*** liegen also in einer (realen) Ebene; die komplanaren realen Geraden **12** $\neq$ **1*2*** definieren den gesuchten Punkt (III IV).

Schliesslich ist es auch leicht, die Gültigkeit der übrigen Veblenschen Axiome festzustellen. Die Geschlossenheitsannahme $\overline{E4}$. stellen wir fest, indem wir einen Dreiraum durch vier nicht in einer Ebene liegende reale Punkte $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$ angeben. Ein beliebiger Punkt $A$ fällt entweder mit $\mathbf{P}_4$ zusammen oder seine Verbindungsgerade mit $\mathbf{P}_4$ trifft die Ebene $\mathbf{P}_1\mathbf{P}_2\mathbf{P}_3$ in einem (idealen) Punkt; $P$ gehört also dem Dreiraum $\mathbf{P}_1\mathbf{P}_2\mathbf{P}_3\mathbf{P}_4$ an. Das Bestehen der Ausdehnungsannahmen E0., E1., E2., E3. liegt auf der Hand. Damit ist eine vollständige Grundlage der projektiven Geometrie gegeben.

Bezüglich einer Definition der "idealen Ebenen," welche wir natürlicher erscheint als die Anwendung der Veblenschen Definition der Ebene, vgl. Anhang.

Wir sehen also: *Ein Raum $\mathfrak{R}$, in welchem Hilbert's Verknüpfungsaxiome und die Annahmen* E), H1), H2) *gelten — lauter Annahmen, welche Beobachtungen in einem beliebigen Gebiet unseres Raumes entsprechen — lässt sich zu einem projektiven Raum ergänzen. Diese Ergänzung kann auf verschiedene Art erfolgen, doch ist unter den in Betracht kommenden algebraischen Körpern einer als der kleinste ausgezeichnet; er ist in allen enthalten. In der Tat ist keines der von uns eingeführten idealen Elemente entbehrlich.*

## Kapitel IV. Reduktion unserer Hypothesen auf schwächere Voraussetzungen

Nun werden wir zeigen, dass sich die einschneidenden Hypothesen H1), H2) durch bedeutend schwächere Annahmen ersetzen lassen, welche mit $\mathfrak{H}'$ (vgl. I.4.) zusammen die Einführung idealer Elemente ermöglichen. Es handelt sich dabei wieder um empirisch naheliegende Annahmen über das Ineinanderliegen von Punkten, Geraden und Ebenen, welche jeden fells im (vollen) projektiven Raume $\mathfrak{P}$ über einem beliebigen Körper $\mathfrak{R}$ zutreffen. Sie haben aber nicht mehr den Charakter notwendiger Bedingungen für die Vervollständigung zu $\mathfrak{P}$. (s. Anhang) und sind formal-logisch etwas kompliziert.

### IV.1. "Raumstück," "Ebenenstück," "Umschliessung" und die Annahmen U), U*), U**)

Ist $A$ ein Punkt ausserhalb der Ebene $\alpha$, so sagen wir: *Der Punkt $P$ gehört dem "Raumstück" $A - \alpha$ an, wenn im Falle $P \neq A$ die Gerade $AP$ $\alpha$ trifft.*

Wir machen nun die folgende Annahme:

U) *Geht durch den Punkt $O$ der Geraden $g$ eine Ebene $\alpha$ hindurch und ist $A$ ein Punkt ausserhalb $\alpha$, so gibt es auf $g$ einen Punkt $P \neq O$ im Raumstück $A - \alpha$.*

Von Interesse ist nur der Fall, dass $A$ auch ausserhalb $g$ liegt, $AP$ trifft dann $\alpha$ in einem Punkte $L \neq O$. Dieser liegt auf der Schnittgeraden a der Ebene $Ag$ mit $\alpha$.

Wir geben dieser Hypothese noch eine andere Fassung. Wir sagen: *Der Punkt $P$ ist "mit Rücksicht auf" $A$, $a$ gewählt, wenn $P \neq A$ ist und die Gerade $AP$ die Gerade $a$ trifft.*

(Falls $A$ ausserhalb $a$ liegt, können wir von dem "Ebenenstück" $A - a$ sprechen.)

Wir machen nun die Annahme:

U!) *Geht durch den Punkt O der Geraden g noch eine Gerade a hindurch und ist A ein Punkt ausserhalb g, überdies a in der Ebene Ag gelegen; so kann man auf g einen Punkt $P \neq O$ mit Rücksicht auf A, a wählen.*

($AP$ trifft dann $a$ in einem Punkte $L \neq O$.)

Von der Annahme U) bezw. U!) machen wir im Anhang zu IV.2. Gebrauch, um H2) auf H1) zurückzuführen.

Im Folgenden (IV.3.) werden wir aber umgekehrt H1) auf H2) zurückführen, mit Hilfe einer Verschärfung von U):

U*) *Gehen durch den Punkt O der Geraden g zwei Ebenen $\alpha$ und $\beta$ hindurch und liegt A ausserhalb $\alpha$, B ausserhalb $\beta$; so gibt es auf g einen Punkt $P \neq O$, welcher zugleich im Raumstück $A - \alpha$ und im Raumstück $B - \beta$ gelegen ist.*

Auch dieser Hypothese geben wir noch eine andere Fassung:

U*!) *Gehen durch den Punkt O der Geraden g zwei weitere Geraden a und b hindurch, und liegen A und B ausserhalb g, ist ferner a in der Ebene Ag, b in der Ebene Bg enthalten; so kann man auf g den Punkt $P \neq O$ zugleich mit Rücksicht auf A, a und mit Rücksicht auf B, b wählen.*

($AP$ trifft dann $a$ in einem Punkte $L \neq O$, $BP$ trifft $b$ in einem Punkte $M \neq O$.)

Schliesslich werden wir aber zum Beweise von H2) in IV.2. eine weitere Verschärfung unserer Hypothese brauchen:

U**) *$\epsilon$ sei eine Ebene welche nicht durch den Punkt O der Geraden g läuft; durch O mögen ferner drei Ebenen $\alpha, \beta, \gamma$ hindurchgehen und es liege A ausserhalb $\alpha$, B ausserhalb $\beta$, C ausserhalb $\gamma$ (A, B, C, in $\epsilon$); dann gibt es auf g einen Punkt $P \neq O$ ausserhalb $\epsilon$, welcher zugleich den Raumstücken $A - \alpha$, $B - \beta$, $C - \gamma$ angehört.*

In anderer Fassung:

U**!) *$\epsilon$ sei eine Ebene, welche nicht durch den Punkt O der Geraden g läuft; durch O mögen ausserdem drei Geraden a, b, c hindurchgehen, die Punkte A, B, C mögen ausserhalb g liegen (A, B, C in $\epsilon$) und a in der Ebene Ag, b in der Ebene Bg, c in der Ebene Cg enthalten sein; dann kann man auf g einen Punkt $P \neq O$ ausserhalb $\epsilon$ zugleich mit Rücksicht auf A, a und auf B, b und auf C, c wählen.*

($AP$ trifft $\alpha$ in einem Punkte $L \neq O$, $BP$ trifft $\beta$ in einem Punkte $M \neq O$, $CP$ trifft $\gamma$ in einem Punkte $N \neq O$.)

Wir geben noch folgende Erklärung: Hat man drei Raumstücke und schneiden sich die definierenden Ebenen in einem Punkte $O$, während die definierenden Punkte in einer Ebene $\epsilon$ liegen, welche nicht durch $O$ geht, so nennen wir den Durchschnitt der drei Raumstücke und des Komplements von $\epsilon$ eine "Umschliessung" von O. Dann besagt unsere Annahme U**):

*Eine Gerade durch O hat mit einer Umschliessung von O einen von O verschiedenen Punkt gemein.*

Diese Annahme ist für den stetigen Raum anschaulich plausibel. Man kann sich vorstellen, dass es nur darauf ankommt, $P$ nicht zu weit weg von $O$ zu wählen.

## IV.2. **Beweis des Desarguesschen Satzes H2) mit Hilfe von U**)**

Wir geben jetzt einen Beweis des Desarguesschen Satzes im Bündel $O$ auf Grund des Axiomensystems $\mathfrak{S}'$ und der Annahme U**). Zunächst formulieren wir den Satz folgendermassen (vgl. II.3.):

*$i, j, h, k, l$ seien Indices, welche die Werte* 1, 2, 3, 4, 5 *durchlaufen. Die zehn Ebenen durch $O$ $(ij) = (ji)$ mögen den Paaren $i \neq j$ entsprechen, die zehn Geraden durch $O$ $(hkl)$ mögen den ungeordneten Tripeln verschiedener Indices entsprechen. Nun wird angenommen:*

1.) (512) *liegt nicht in* (14),
2.) (123) *liegt nicht in* (15),
3.) (234) *liegt weder in* (12) *noch in* (13).

*Behauptet wird*: (345) *liegt in* (45), *falls alle übrigen Inzidenzen von $(ijk)$ mit $(jk)$ stattfinden.*

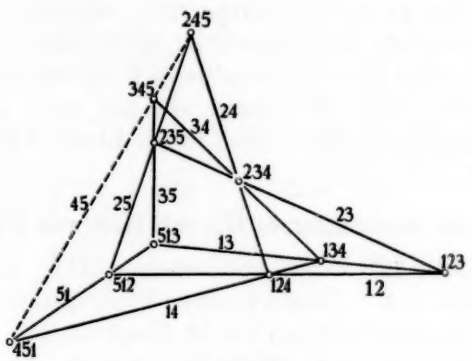



FIG. 5

Der Beweis kann nun folgendermassen geführt werden:

1. Wähle den Punkt [451] $\neq O$ auf (451).
2. Wähle den Punkt [512] $\neq O$ auf (512) "mit Rücksicht auf" [451] und (513). So erhalten wir den Punkt [513] $\neq O$ auf (513) und die Gerade [51], welche mit den drei Punkten inzidiert; sie geht nicht durch $O$. (1.))
3. Wähle [123] $\neq O$ auf (123) "mit Rücksicht auf" [512], (124) und "mit Rücksicht auf" [513], (134) zugleich. So erhalten wir [124] $\neq O$ auf (124) sowie [12] durch die Punkte [123], [124], [512] und [134] $\neq O$ auf (134) sowie [13] durch [132], [134], [135]. Alle sechs bisher gewählten Punkte liegen in einer nicht durch $O$ gehenden Ebene [1]. (2.)) Sie verbindet [123] mit [15]. In ihr liegt auch die Gerade [14] als Schnitt mit (14). Sie geht durch [142], [143], [145].
4. Wähle [234] $\neq O$ auf (234) ausserhalb [1] zugleich
   "mit Rücksicht auf" [123], (235),
   "mit Rücksicht auf" [124], (245), und
   "mit Rücksicht auf" [134], (345).

So erhalten wir [235] $\neq O$ auf (235) sowie [23] durch [231], [234], [235]. Durch diese Gerade gehen die beiden Ebenen [2] und [3], welche [234] mit [12] bezw. [13] verbinden. Sie gehen nicht durch $O$. (3.))

Ferner erhalten wir [245] $\neq O$ auf (245) sowie [24] durch [243], [241] und [245]. Wir erhalten noch [345] $\neq O$ auf (345) sowie [34] durch [342], [341], [345].

Die Ebene [2] schneidet (25) in einer Geraden [25], welche [251], [253], [254] enthält.

Die Ebene [3] schneidet (35) ineiner Geraden [35], welche [351], [352], [354] enthält.

Ferner erhalten wir eine Ebene [4] indem wir [234] mit [14] verbinden (sie kann durch $O$ gehen). In ihr liegen [24] und [34], also [245] und [345], aber auch [145].

[235] kann nicht in [15] liegen, sonst lägen [23] und [234] in [1] gegen 4, oder wäre [235] = [123] gegen 2.1.

Wir erhalten also eine Ebene [5], welche [235] mit [15] verbindet. In ihr liegen mit [25], [35] gleichfalls die Punkte [451], [452], [453].

Es ist aber [5] $\neq$ [4], sonst lägen alle erhaltenen Punkte in der Ebene [1].

Die drei Punkte [451], [452], [453] liegen also auf einer Geraden [45], der Schnittgeraden von [4] und [5]; daher (453) in der Ebene (45), welche [451] $\neq$ [452], also [45] enthält.

### IV.3. Beweis des Vierkantsatzes H1) mit Hilfe von H2) und U*)

Nun setzen wir $\mathfrak{H}'$, H2), U*) voraus und beweisen H1). Es sei ein Dreikant gegeben, dessen Ebenen 3, 4, 5 keinen (realen) Punkt gemein haben. 4 und 5 schneiden sich in der Geraden 45, 5 und 3 in 53, 3 und 4 in 34. $g$ sei eine Gerade, welche mit 53 und 34 komplanar ist und nicht in 3 liegt. Zu beweisen ist dann die Komplanarität von $g$ mit 45. Dies kann so geschehen:

Wähle den Punkt 145 auf 45, den Punkt 153 auf 53, den Punkt 134 auf 34. Diese drei Punkte liegen nicht auf einer Geraden, ihre Verbindungsebene heisse 1. Sie schneidet 3 in 13, 4 in 14, 5 in 15.

Sodann wähle den Punkt 245 $\neq$ 145 auf 45; ferner 253 $\neq$ 153 auf 53 "mit Rücksicht auf" 245, 15. (53 und 15 gehen durch 153; 15 liegt in der Ebene 5, welche 245 mit 53 verbindet.) So erhalten wir den Schnittpunkt 125 und die Gerade 25 durch 251, 253, 254. (U!) mit 153 als Punkt $O$.)

Schliesslich wähle 234 $\neq$ 134 auf 34 zugleich "mit Rücksicht auf" 245, 14 (in 4) und "mit Rücksicht auf" 235, 13 (in 3). So erhalten wir den Schnittpunkt 124 und die Gerade 24 durch 241, 243, 245, sowie den Schnittpunkt 123 und die Gerade 23 durch 231, 234, 235. Die Punkte 245, 253, 234 liegen nicht auf einer Geraden, ihre Verbindungsebene heisse 2. Sie schneidet 1 in einer Geraden 12, auf welcher die Punkte 123, 124, 125 liegen. (U*!) mit 134 als Punkt $O$.)

Das Dreieck 145, 153, 134 und das Dreieck 245, 253, 234 sind perspektiv in Bezug auf die Gerade 12. Von den Ebenen 1, 2, 3, 4, 5 gehen keine vier durch einen Punkt. $g$ liegt weder in 1 noch in 2. (Eine nicht durch 134 gehende Gerade von 1 muss nämlich zu der nicht in 1 liegenden Geraden 34 windschief

sein. Läge also $g$ in 1, so müsste $g$ durch 134 und ebenso auch durch 135 gehen. Dann läge $g$ aber in 3 gegen die Voraussetzung.)

Nun können wir auf $g$ einen Punkt $O$ ausserhalb der Ebenen 1 und 2 wählen. (E)) $O$ liegt auch ausserhalb 3. In der Projektion von $O$ aus sind jetzt die Voraussetzungen von H2) gegeben. Wir bezeichnen die Projektion von $ij$ mit $(ij)$, die Projektion von $hkl$ mit $(hkl)$. Die Ebenen (34) und (35) schneiden sich in (345) = $g$. H2) lehrt, dass (345) in (45) liegt, d.h. dass $g$ mit (45) komplanar ist, was zu beweisen war.

Damit ist die ganze Theorie auf Hilbert's Verknüpfungsaxiome und unsere Umschliessungsannahmen zurückgeführt.

Es scheint mir bemerkenswert, dass man bereits mit diesen Annahmen durchkommt. Ob eine weitere Reduktion möglich ist, bleibe dahin gestellt. (In vier Dimensionen käme man bereits mit dem Analogon von U*) aus.)

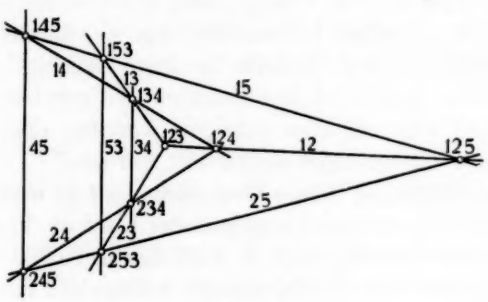



FIG. 6

## ANHANG

### Einige Bemerkungen zur Geschichte des Gegenstandes

Chr. v. Staudt in seiner "Geometrie der Lage" (Nürnberg 1841) war wohl der erste, der die projektive Geometrie systematisch als Theorie des Ineinanderliegens von Punkten, Geraden und Ebenen entwickelt hat. Er abstrahiert die grundlegenden Aussagen aus der Euklidischen Geometrie und zieht namentlich das Parallelenaxiom zur Einführung der unendlich fernen Elemente heran. In seinen "Beiträgen zur Geometrie der Lage" (Nürnberg 1857) entwickelte er einen geometrischen Kalkül, die "Wurfrechnung," welche auf seinem Fundamentalsatz beruht. In der Begründung spielen Stetigkeitsbetrachtungen eine Rolle.

Die Möglichkeit, die projektive Geometrie auf Axiome zurückzuführen, deren empirische Grundlage Beobachtungen in einem beschränkten Raumstück bilden, hat zuerst Felix Klein erkannt. Nachdem er—anknüpfend an eine Arbeit Cayley's—auf die Isomorphie des Lobatschefskijschen Raumes mit einem Teilgebiet des (reellen) projektiven Raumes aufmerksam geworden war, überlegte er sich, dass Staudt's wichtigste Konstruktionen bereits in einem beschränkten Raumstück durchgeführt werden können und skizzierte eine

Theorie der Einführung idealer Elemente. (1871/72 "Über die sogenannte Nicht-Euklidische Geometrie"; Math. Annalen 4, S. 623 f., namentlich §17, "Die Unabhängigkeit der projektivischen Geometrie von der Parallelentheorie"; 6, S. 134 ff.)

Er sagt: "In einem begrenzten Raume sei eine unendliche Zahl überall stetig gekrümmter, nur durch die Begrenzung des Raumes geendeter Flächen gegeben, welche die folgende Gruppierung besitzen:

1.) Durch drei beliebig angenommene Punkte des gegebenen Raumes geht eine [und nur eine] Fläche des Systems hindurch.

2.) Die Durchschnittskurve, welche zwei Flächen des Systems gemein haben können, gehört allen Flächen an, die zwei Punkte der Kurve enthalten."

(Die von mir in eckige Klammern eingeschlossenen Worte in 1.) bedeuten wohl "im allgemeinen" nur eine. Sie können jedoch ganz weggelassen werden.)

"Für ein solches System von Flächen und Kurven gilt die projektivische Geometrie in demselben Sinne wie gemäss den gewöhnlichen Vorstellungen für das System der Ebenen und Geraden in einem beliebig begrenzten Raume. Anders ausgesprochen: Man wird den Punkten des gegebenen Raumes in der Art Zahlen zuordnen (Koordinaten erteilen) können, dass die Flächen des Systems durch lineare Gleichungen dargestellt werden."

Die genauere Durchführung dieses Gedankens auf Grund von ausdrücklich ausgesprochenen Axiomen verdankt man in erster Linie M. Pasch's "Vorlesungen über neuere Geometrie" (Leipzig 1882, 2. Aufl. Berlin 1926). Pasch kam es vor allem darauf an, alle für einen empiristischen Aufbau der Geometrie geeigneten Grundbegriffe und Grundtatsachen aufzuspüren. Die begrenzte Strecke und das begrenzte Ebenestück erschienen ihm als einzig möglicher Ausgangspunkt.

Hilbert stellte sich um die Jahrhundertwende in seiner Festschrift "Grundlagen der Geometrie" die Aufgabe, die einfachsten Axiome auszuwählen, welche eine vollständige Grundlage für den logischen Aufbau der Geometrie bilden, ohne den Gesichtspunkt der anschaulichen Natürlichkeit ausser Acht zu lassen. Er führte die "Strecke" als Punktepaar ein und ersetzte das "Ebenenstück" durch "drei Punkte, welche nicht auf einer Geraden liegen." So konnte er die "Gerade" als gemeinsame Marke (Klasse) gewisser Punktepaare, die "Ebene" als gemeinsame Marke gewisser Dreiecke auffassen. Er sonderte die Verknüpfungsaxiome von den Anordnungsaxiomen, doch interessierte ihn damals die Tragweite der ersteren allein nicht weiter. Auch in Kap. V, wo er auf Grund der Verknüpfungsaxiome und des Parallelenaxioms zum Desarguesschen Satz und der auf ihn beruhenden Streckenrechnung gelangt, führt er die Anordnungsaxiome mit an, wohl mit Rücksicht auf die Anordnungs-Konsequenzen. Sein "Desarguessches Zahlensystem" ist nichts anderes als was wir heute einen geordneten Schiefkörper oder kürzer geordneten Körper nennen.

Aus Pasch's Vorlesungen geht bereits hervor, dass die Axiome der Verknüpfung und der Anordnung hinreichen, um die Theorie der idealen Elemente zu begründen. Das durch die ergänzende und vereinfachende Arbeit einer Reihe von Geometern schliesslich erzielte Ergebnis fand seine Darstellung durch A. N.

Whitehead in Cambridge Tracts No. 4 (*The Axioms of Projective Geometry*), No. 5 (*The Axioms of Descriptive Geometry*), 1906/7, repr. 1913/14. Die Einführung der Anordnung im deskriptiven Raume sucht W. mit den Worten zu motivieren: "· · · the purely classificatory portions of a Descriptive Geometry are clumsy and uninteresting." Durch die vorliegende Arbeit wird—wie ich glaube—diese Auffassung widerlegt.

## Ergänzungen und Literaturhinweise

### Zu I.1

In Axiom c) sind die Bedingungen des Auseinanderliegens von $P_i$ und $\epsilon_k$, wenn $i - k \equiv -2 \pmod 5$ von den übrigen Bedingungen abhängig. Zunächst erkennen wir, dass vier der Punkte z.B. $P_1$, $P_2$, $P_3$, $P_4$ nicht in einer Ebene $\epsilon$ liegen können. ($P_1$, $P_2$, $P_3$ liegen in $\epsilon_2$. Es wäre $\epsilon_2 \neq \epsilon$, weil $P_4$ ausserhalb $\epsilon_2$ liegt. $P_1$, $P_2$ liegen ausserdem in $\epsilon_1$. Nach Axiom a) müsste $P_3$ in $\epsilon_1$ liegen gegen die Voraussetzung.) Nun enthält $\epsilon_k$ die Punkte $P_{k-1}$, $P_k$, $P_{k+1}$; enthielte $\epsilon_k$ ausserdem $P_{k-2}$, so lägen vier der Punkte in einer Ebene. (Die Indices gelten mod 5).

### Zu I.4

Zu B2*1): Die gemeinsame Gerade geht durch A hindurch. (wegen A1))

Zu C1): Auf Grund von B2*1) schneiden sich $\epsilon_i$ und $\epsilon_k$ ($i \neq k$) in einer Geraden $g_{ik}$, welche weder durch $P_i$ noch durch $P_k$ läuft, aber durch die beiden anderen Punkte. $P_1$, $P_2$, $P_3$, $P_4$ liegen nicht in einer Ebene, auch gehen $\epsilon_1$, $\epsilon_2$, $\epsilon_3$, $\epsilon_4$ nicht durch einen Punkt.

Beweis von A1*): $a$ liege nicht in $\alpha$ und habe mit $\alpha$ die Punkte $A \neq B$ gemein. Man kann durch $a$ eine Ebene $\beta$ legen (C1), B1)), sie ist $\neq \alpha$ und hat mit $\alpha$ $A$ und $B$ gemein. Die Schnittgerade $b$ (B2*1)) ist von a verschieden und geht durch $A$ hindurch. (A1)), ebenso durch $B$. Dies widerspricht A2).

Beweis von A2*): Hätten $\alpha$ und $\beta$ zwei Geraden $a \neq b$ gemein, so gäbe es auf $b$ zwei Punkte (B1*2)), von denen mindestens einer $B$ ausserhalb a läge (A2)). Auch $B$ läge in $\alpha$ und $\beta$ (A0)). Dies widerspräche A1).

Beweis von B2): C1) liefert uns eine Ebene $\epsilon$, welche nicht durch $A$ geht. In $\epsilon$ liegt ein Dreieck. $A$ kann mit den Seiten desselben durch Ebenen verbunden werden (B1)). Diese schneiden sich paarweise in Geraden (B2*1)). Diese Geraden geben—falls nicht schon eine von ihnen durch $B$ geht—mit $B$ verbunden drei Ebenen, von denen sicher zwei verschieden sind. (Bedingung für $\epsilon$) Beide enthalten $A$ und $B$. Ihre Schnittgerade (B2*1)) enthält $A$ und $B$. (A1))

### $\mathfrak{H}$ besteht aus unabhängigen Axiomen

Die Unabhängigkeit von A0) zeigen wir, indem wir die Punkte und Ebenen einer projektiven Geometrie durch eine einzige Gerade ergänzen, welche mit allen Punkten und allen Ebenen inzidiert.

Die Unabhängigkeit von A1) zeigen wir, ausgehend von einer projektiven Geometrie mit Hilfe einer Doppelgängerin einer Ebene, d.h. einer weiteren Ebene, welche einer bestimmten Ebene in allen Inzidenzen folgt.

Die Unabhängigkeit von A2) zeigen wir analog mit Hilfe einer Doppelgängerin einer Geraden.

Um die Unabhängigkeit von B1) zu zeigen, schliessen wir eine Ebene aus einer projektiven Geometrie aus aber unter Beibehaltung der mit ihr inzidierenden Elemente.

Um die Unabhängigkeit von B1*1) zu zeigen, fügen wir eine unverknüpfte Ebene zu einem affinen Raum hinzu.

Um die Unabhängigkeit von B1*2) zu zeigen, fügen wir zu einem affinen Raum lediglich eine (unendlich ferne) Gerade hinzu, welche nur einer Schaar paralleler Ebenen angehört.

Die Unabhängigkeit von B2*1) zeigt ein vierdimensionaler Raum.

C1) ist die einzige Forderungen, welche direkt die Existenz von Elementen verlangt.

Mann kann—wenn man in B1) die eingeklammerte Bedingung beachtet, —C1) noch in zwei unabhängige Annahmen spalten:

C1.1) *Es gibt einen Punkt und eine Gerade, welcher nicht miteinander inzidieren.*

C1.2) *Es gibt zwei Geraden, welche keine Ebene gemein haben.*

Die Unabhängigkeit von C1.1) zeigt eine Gerade mit zwei Punkten auf ihr. (Lineare Geometrie)

Die Unabhängigkeit von C1.2) zeight eine Ebene, in der ein Dreieck liegt. (Ebene Geometrie)

Die Unabhängigkeit von A0) muss in diesem Falle anders gezeigt werden, etwa durch ein Tetraeder mit seinen Ecken, Kanten und Flächen und eine Ebene $\epsilon$ durch eine Ecke, welche mit den Seiten des gegenüberliegenden Dreiecks inzidiert, nicht aber mit seinen Ecken.

*Hilbert's Axiome der Verknüpfung I* lauten:

1. *Zu zwei Punkten $A \neq B$ gibt es stets eine Gerade $a$, die mit jedem der beiden Punkte $A$, $B$ zusammengehört.*
2. *Zu zwei Punkten $A \neq B$ gibt es nicht mehr als eine Gerade $a$, die mit jedem der beiden Punkte $A$, $B$ zusammengehört.*

3a. *Auf einer Geraden gibt es stets wenigstens zwei Punkte.*

3b. *Es gibt wenigstens drei Punkte, die nicht auf einer Geraden liegen.*

4a. *Zu irgend drei nicht auf ein und derselben Geraden liegenden Punkten, $A$, $B$, $C$ gibt es stets eine Ebene $\alpha$, die mit jedem der drei Punkte $A$, $B$, $C$ zusammengehört.*

4b. *Zu jeder Ebene gibt es stets einen mit ihr zusammengehörigen Punkt.*

5. *Zu irgend drei nicht auf ein und derselben Geraden liegenden Punkten $A$, $B$, $C$ gibt es nicht mehr als eine Ebene, die mit jedem der drei Punkte $A$, $B$, $C$ zusammengehört.*
6. *Wenn zwei Punkte $A \neq B$ einer Geraden $a$ in einer Ebene $\alpha$ liegen, so liegt jeder Punkt von $a$ in der Ebene $\alpha$.*

7. *Wenn zwei Ebenen $\alpha \neq \beta$ einen Punkt A gemein haben, so haben sie wenigstens noch einen weiteren Punkt B gemein.*
8. *Es gibt wenigstens vier nicht in einer Ebene gelegene Punkte.*

Diese Axiome sind unabhängig:

Eine Figur, welche die Unabhängigkeit von 1. zeigt, bilden die Ecken und Flächen eines Tetraeders allein.

Um die Unabhängigkeit von 6. zu zeigen, gehe man von einem Tetraeder aus und identifiziere in einer Seitenfläche die Geraden.

Für die übrigen Unabhängigkeitsbeweise ergeben sich Beispiele aus dem Vorhergehenden. Für später (III.2.) sei noch angemerkt, dass aus Hilbert's Verknüpfungsaxiomen folgt:

In einer Ebene gibt es stets wenigstens drei nicht auf einer Geraden gelegene Punkte. (Rosenthal, Math. Annalen 69)

Die Gleichberechtigung der Inzidenz von Gerade und Ebene mit den übrigen Inzidenzbeziehungen wurde auch von Nil Glagoleff geltend gemacht. Dieser gibt in Comptes Rendus 201 p. 867 f. (Paris 1935) ein in ähnlicher Weise formuliertes Axiomensystem, welches die Axiome der Verknüpfung und das Parallelenaxiom zusammenfasst. Er hat nicht alle Abhängigkeiten bemerkt, z.B. nicht die von B2).

## Zu I.5

*$\mathfrak{A}$ besteht aus unabhängigen Axiomen.*

Die Unabhängigkeit der Axiome A0), A1), A2), B1) wird durch dieselben Beispiele gezeigt, wie im vorhergehenden Abschnitt. Die Unabhängigkeit von B1*) ergibt sich durch Ausschluss eines Punktes aus einem projektiven Raum.

Die Unabhängigkeit von B2*) ergibt sich durch Ausschluss einer Geraden aus einem projektiven Raum.

Die Unabhängigkeit von C) ist evident; seine Nichtersetzbarkeit durch C1) zeigt das Tetraeder mit seinen Ecken, Kanten und Flächen.

Der Erste, der ein Axiomensystem der projektiven Geometrie so ausgesprochen hat, dass die Punkte, Geraden und Ebenen als gleichberechtigte Elemente auftreten, ist wohl Enriques in seinen "Vorlesungen über projektive Geometrie" (Deutsche Ausgabe 2. Aufl. Teubner 1915). Dort wird die projektive Geometrie auf die Existenz und Eindeutigkeit der grundlegenden Verknüpfungen von zwei und drei Elementen zurückgeführt. Die Annahmen sind nicht unabhängig.

Neuerdings kam N. Glagoleff in Periodico di Matematiche (4)15, p. 172/76 (Bologna 1936) darauf zurück. Er gibt ein einfacheres Axiomensystem an, welches in unserer Bezeichnungsweise aus den Axiomen besteht: A0), A1) und B1), A2) und B2), B1*), A2*) und B2*), ausserdem die Annahme: *Es gibt vier Punkte, die weder auf einer Geraden noch in einer Ebene gelegen sind.* (Die letzte Annahme kommt auf C1) hinaus und ist zu schwach, wenn man die Geometrie auf einen Körper zurückführen will.)

Glagoleff wirft Enriques nicht ganz mit Recht vor, das von uns mit A0)

bezeichnete Axiom folge nicht aus seinen Annahmen. Aber man kann A0) beweisen, wenn man die Annahmen macht: *Auf jeder Geraden liegen mindestens zwei Punkte; in einer Ebene ausserhalb einer ihrer Geraden liegt ein Punkt.*

Liegt $A$ auf $a$, $a$ in $\alpha$ und ist $B$ ein zweiter Punkt auf $a$, $C$ ein Punkt in $\alpha$ ausserhalb $a$, so wird die Ebene $ABC$ $a$ enthalten. (A1*)), also gemäss A1) mit der Ebene $\alpha$ identisch sein, welche daher durch $A$ hindurchgeht.

(Gl.'s Beispiel für die Unabhängigkeit von A0) ist unzutreffend, es kann im Rahmen seiner Annahmen durch folgendes ersetzt werden: Ein Tetraeder und eine Gerade, welche durch eine Ecke geht und in der gegenüberliegenden Fläche liegt.)

Ein vollständiges Axiomensystem der projektiven Geometrie, in welchem Punkte, Geraden und Ebenen als gleichberechtigte Elemente auftreten gibt auch Bieberbach, der in seiner "Einleitung in die höhere Geometrie" (Leipzig 1933) glaubte, der erste zu sein, welcher ein derartiges, zugleich sich selbst duales System aufstellt. Ein unabhängiges Axiomensystem erzielt er nur durch geeignete Zusammenfassung. Seine Axiome sind in unserer Bezeichnungsweise: A0), A2) und B2), A2*) und B2*), A1*), (A1)), G), G*), E), E*), ferner Veblen's E1., E3., E3*. (dual zu E3.)

## Zu II.2

Ein endliches Beispiel für dieselbe Behauptung ergibt sich ähnlich, wenn man von dem aus vier Elementen bestehenden Körper $\mathfrak{R}(\epsilon)$ ausgeht, der durch das Rechnen mod $(2, \epsilon^2 + \epsilon + 1)$ erhalten wird, und sich der folgenden Multiplikationsvorschrift für die Zahlenpaare bedient:

$$\eta_1 = \alpha_1 \xi_1 + \epsilon \alpha_2^2 \xi_2,$$

$$\eta_2 = \alpha_2 \xi_1 + \alpha_1^2 \xi_2.$$

Hier versagt das Assoziativgesetz.

Dass Hilbert's Verknüpfungsaxiome allein nicht genügen, um den Raum zu einem Teil eines projektiven Raumes zu machen, also auf einen Körper zurückzuführen, zeige ich am einfachsten so:

Man denke sich eine Nicht-Desarguessche Ebene, in welcher ausser den ebenen Verknüpfungsaxiomen (Hilbert I.1–3) das Parallelenaxiom in der schärferen Fassung IV* ("eine und nur eine") erfüllt ist, in zwei Exemplaren $\epsilon$, $\epsilon'$. Diese Ebenen seien so übereinander gelegt, dass man zu einer Geraden in der einen durch einen Punkt in der anderen eine und nur eine Parallele ziehen kann. Ein solches Paar paralleler Geraden heisse "Ebene," ferner $\epsilon$ und $\epsilon'$. "Punkt" und "Inzidenz" haben ihre gewöhnliche Bedeutung. Dann sind die Verknüpfungsaxiome $\mathfrak{h}$ erfüllt. Die Ungültigkeit der Vierkanthypothese zeigt sich wie in II.2. (Übrigens auch die von H2))

## Zu III.2

Die Art der Einführung idealer Geraden erinnert an Pasch.

## Zu III.3

Die *Erklärung der "idealen Ebenen"* kann vollkommen dual zur Definition der "idealen Punkte" in III.1. gegeben werden: Sind $A$, $B$, $C$ drei nicht auf einer Geraden gelegene Punkte, so sagen wir: "Die Gerade $g$ schliesst sich dem Dreieck $ABC$ an," wenn sie alle drei Seiten desselben trifft. Schliessen sich zwei Geraden demselben Dreieck an, so haben sie einen Punkt gemein. Wenn sich die Seiten eines Dreiecks $A'B'C'$ den Dreieck $ABC$ anschliessen, so nennen wir die beiden Dreiecke "ebenengleich." Diese Beziehung ist symmetrisch und transitiv. (A3.) Die Klasse aller zu dem Dreieck $ABC$ ebenengleichen Dreiecke bildet die "Ebene" $ABC$. Wir sagen: "Die Gerade $g$ liegt in der Ebene $ABC$," wenn $g$ sich dem Dreieck $ABC$ anschliesst; "Ein Punkt liegt in der Ebene $ABC$," wenn er auf einer solchen Geraden $g$ liegt.

Ein interessantes *Beispiel zur Einführung idealer Elemente* liefert die ganzzahlige Affingeometrie, d.h. die Betrachtung eines räumlichen Gitters mit den Gittererbenen und Gittergeraden. Sie führt auf den Körper der rationalen Zahlen.

Zu beachten ist auch das Beispiel, das sich aus unserer Konstruktion in II.2. ergibt, wenn die dort auftretende Funktion $f$ linear ist. (Drei- und vierdimensionale "Ebenen" im fünfdimensionalen Raum) Es führt auf den Körper $\Re(i)$ mit $i^2 = -1$; der zugehörige projektive Raum ist reell-sechsdimensional.

## Zu IV

Dass U) *keine notwendige Bedingung* für die Einführung idealer Elemente ist, zeigt das erste Beispiel in Zusatz zu III.3.

## Zu IV.1

Kehren wir zu der Auffassung zurück, von welcher wir in I.2. ausgegangen sind, bei welcher die "Gerade" als abgeleiteter Begriff aufgefasst wird, so zeigt sich, dass *bei Einführung der Annahme* U) das Axiom b*2) *entbehrlich* wird, wenn man b*1) verschärft zu:

b*1′) *In jeder Ebene liegen drei Punkte, durch welche keine andere Ebene hindurchgeht.*

Hat man nämlich zwei verschiedene Ebenen $\alpha$, $\beta$ durch $O$, so wähle man in $\beta$ eine durch $O$ gehende Gerade $g$ und einen Punkt $A$ ausserhalb $g$. Dann findet man zufolge U) auf $g$ einen Punkt $P \neq O$ derart, dass die Gerade $AP$ die Ebene $\alpha$ in einem Punkt $L \neq O$ trifft. Dieser Punkt liegt offenbar in $\alpha$ und $\beta$ zugleich.

## Zu IV.2

### Zurückführung von H2) auf H1) mit Hilfe von U)

Unter Voraussetzung der Hypothese H1), welche die Einführung idealer Punkte ermöglicht, genügt es dafür zu sorgen, dass je zwei Ebenen $[i]$, $[j]$ sich schneiden. Hiefür reicht es hin, dass die Punkte **[451]**, **[512]**, **[513]** und **[234]**

real sind. Dies erfordert nur, dass **[512]** unter Anwendung von U!) wie im Text angegeben gewählt wird. **[234]** $\neq O$ auf (234) muss nur ausserhalb von [1] gewählt werden (E)).

## Zu IV.3

Die Beweismethode erinnert an eine bei Whitehead l. c. wiedergegebene, wohl von Bonola herrührende Beweisführung.

PRAGUE, CZECHOSLOVAKIA.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# CONCERNING RINGS OF CONTINUOUS FUNCTIONS

BY M. EIDELHEIT

(Received April 10, 1939)

Let $E$ be the ring of all real continuous functions $x(t)$ defined for $t \in T$, where $T$ is a given compact metric space. In $E$ a norm can be defined:

$$|x| = \max_{t \in T} |x(t)|.$$

$E$ is then a separable space of the type $(B)$.[1] S. Mazur, in his researches concerning the rings of this kind,[2] proposed the question, whether *every closed subring $P$ of $E$ containing the unit element $x(t) \equiv 1$ is equivalent to the ring of all continuous functions defined on some compact metric space.* We give here a positive answer for this question.

Let $\{x_n\}$ be a sequence dense in the unit sphere $|x| \leqq 1$ of $P$. Denote by $S$ the set of the sequences $\{\xi_n\}$, where

$$\xi_n = x_n(t) \qquad (n = 1, 2, \cdots ; t \in T).$$

$S$ is a compact and closed subset of the Fréchet space $(s)$.[3] Now let $y(\xi)$, $(\xi = \{\xi_n\})$ be any continuous function defined on $S$. We put for $t \in T$

$$x(t) = y(x_1(t), x_2(t), \cdots ). \tag{1}$$

We have to show that $x(t) \in P$. By a theorem of Tietze[4] there exists a real continuous function $F(\xi)$ defined on the whole space $(s)$ such that

$$F(\xi) = y(\xi) \quad \text{for } \xi \in S.$$

Denote by $K$ the set of points $\xi = \{\xi_n\}$ with $|\xi_n| \leqq 1$ $(n = 1, 2, \cdots )$. $K$ is a compact closed set, hence $F(\xi)$ is uniformly continuous on $K$. On the other hand we have for $\xi = \{\xi_n\}$, $\xi^{(i)} = (\xi_1, \xi_2, \cdots, \xi_i, 0, 0, \cdots )$ uniformly in $K$. It follows then that

$$\lim_{i \to \infty} \xi^{(i)} = \xi$$

$$F(\xi) = \lim_{i \to \infty} F(\xi^{(i)})$$

uniformly in $K$. But $F(\xi^{(i)})$ is a continuous function of $i$ variables $\xi_1$, $\xi_2, \cdots, \xi_i$, hence there exists a sequence of polynomials $\{W_n^{(i)}(\xi_1, \xi_2, \cdots, \xi_i)\}$

[1] K. Borsuk, *Sur les rétractes*, Fund. Math., 17, 1931, p. 153–170, th. 2.

[2] To be published in the "Studia Mathematica."

[3] See: S. Banach, *Théorie des opérations linéaires*, Warszawa 1932, p. 10.

[4] H. Tietze, *Über Funktionen, die auf einer abgeschlossenen Menge stetig sind*, Journ. f. reine u. angew. Math. 145/1915/p. 9–14, Satz 3.

tending uniformly to $F(\xi^{(i)})$ in $K$. For a suitable sequence $\{k_i\}$ we have then

$$F(\xi) = \lim_{i\to\infty} W_{k_i}^{(i)}(\xi_1, \xi_2, \cdots, \xi_i)$$

uniformly in $K$, and since $S \subset K$ we have also

$$y(\xi) = \lim_{i\to\infty} W_{k_i}^{(i)}(\xi_1, \xi_2, \cdots, \xi_i)$$

uniformly in $S$.[5]

$P$ being a subring of $E$ the function $W_{k_i}^{(i)}(x_1(t), x_2(t), \cdots)$ belongs to $P$. Since the sequence $\{W_{k_i}^{(i)}(x_1(t), x_2(t), \cdots)$ converges uniformly to $x(t)$ for $t \in T$ we have by (1) also $x(t) \in P$.

Conversely, let $x(t)$ be any function belonging to $P$. We define then $y(\xi)$ in the following way. If $\xi = \{\xi_n\}$, $\xi \in S$, there exists a point $t \in T$

$$\xi_n = x_n(t) \qquad (n = 1, 2, \cdots);$$

then we put

$$y(\xi) = x(t).$$

We shall prove now that this definition determines $y(\xi)$ uniquely and that $y(\xi)$ is a continuous function on $S$. Let $t_1$ and $t_2$ be two points of $T$ such that

$$x_n(t_1) = x_n(t_2) \qquad (n = 1, 2, \cdots).$$

The sequence $\{x_n(t)\}$ being dense in the unit sphere of $P$, we have then also $x(t_1) = x(t_2)$, hence $y(\xi)$ is uniquely defined.

Now let

$$\lim_{i\to\infty} \xi^{(i)} = \xi; \quad \xi^{(i)} = \{\xi_n^{(i)}\}, \quad \xi = \{\xi_n\}, \quad \xi_n^{(i)} = x_n(t_i), \quad \xi_n = x_n(t) \qquad (t_i, t \in T).$$

If $\{t_{i_k}\}$ is any convergent partial sequence $\lim_{k\to\infty} t_{i_k} = t'$, we have

$$\lim_{k\to\infty} x_n(t_{i_k}) = \lim_{k\to\infty} \xi_n^{(i_k)} = \xi_n = x_n(t) \qquad (n = 1, 2, \cdots),$$

hence, the function $x_n(t)$ being continuous,

$$x_n(t) = x_n(t') \qquad (n = 1, 2, \cdots).$$

It follows then as before that

$$x(t) = x(t').$$

Let now $\{\xi^{(i_k)}\}$ be any subsequence of the sequence $\{\xi^{(i)}\}$. We may suppose that the corresponding sequence $\{t_{i_k}\}$ is convergent. We have then by the former remark

$$\lim_{k\to\infty} y(\xi^{(i_k)}) = \lim_{k\to\infty} x(t_{i_k}) = x(t) = y(\xi),$$

[5] This generalized theorem of Weierstraß was proved for the first time by M. Fréchet.

hence, as the sequence $\{y(\xi^{(i)})\}$ is bounded,

$$\lim_{i\to\infty} y(\xi^{(i)}) = y(\xi).$$

Denote now by $R$ the ring of all real continuous functions on $S$, and put for $x \in P$

$$U(x) = y,$$

where $y = y(\xi)$ is defined as above. We obtain then obviously an isomorphism between $P$ and $R$ which preserves besides the rules of addition and multiplication also the norm.

Lwów.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# LA THÉORIE DU CORPS DE CLASSES

PAR C. CHEVALLEY

(Received December 14, 1939)

## Introduction

La théorie du corps de classes se présente un peu plus simplement aujourd'hui qu'il y a quelques années, notamment du fait de l'élimination des "moyens transcendants." Cependant l'on y sent encore bien des complications, probablement superflues; le présent mémoire aurait donc trouvé la seule justification de son existence si sa lecture induisait de nouveaux chercheurs à poursuivre le travail de simplification et surtout de clarification ici entrepris.

La forme dans laquelle sont enoncés les théorèmes relatifs aux corps de classes (nous les avons condensés en un seul énoncé, §5, p. 404) risque de surprendre au premier abord ceux des lecteurs qui connaissent déjà quelque peu la théorie. Quelques mots d'explication sont donc nécessaires.

Tout d'abord, nous avons substitué au langage classique de la théorie des idéaux l'emploi d'une notion nouvelle, celle de l'*idèle*. Cette substitution était nécessaire pour inclure dans la théorie le cas des extensions abéliennes infinies; mais elle comporte déjà certains avantages dans le cas classique, toutes les fois qu'interviennent des ramifications; elle permet en effet d'éviter le maniement toujours un peu délicat des "groupes de congruence,"[1] avec leurs multiples "modules de définition."

Nous avons déjà défini les idèles dans un mémoire antérieur,[2] dans lequel se trouve expliqué le mécanisme de passage des groupes de congruence classiques aux groupes d'idèles. Nous redonnons ici la definition *in abstracto* des idèles, renvoyant au mémoire en question pour ce qui est de la comparaison avec les groupes de congruence.

L'objet de la théorie du corps de classes est de montrer comment les extensions abéliennes d'un corps de nombres algébriques $K$ peuvent être déterminées par des éléments tirés de la connaissance de $K$ lui-même; ou, si l'on veut présenter les choses en termes dialectiques, comment un corps possède en soi les éléments de son propre dépassement (et ce, sans aucune contradiction interne!).

Au lieu de penser la totalité des extensions abéliennes de $K$, il est peut-être plus simple de se référer à l'extension abélienne maxima de $K$, dans laquelle

[1] Voir mon précédent travail *Sur la théorie du corps de classes dans les corps finis et les corps locaux*, Journ. of Coll. of Sciences, Tokyo, 1933, II, 9, p. 365. Ce mémoire sera designé par *Thèse* dans les références ultérieures.

[2] *Généralisation de la théorie du corps de classes pour les extensions infinies.* Liouville, IX, vol. 15, 1936, p. 359.

toutes les autres sont contenues: c'est ce point de vue que nous adopterons de préférence au cours du présent travail. Or, il se trouve que l'on peut construire le groupe de Galois de cette extension abélienne maxima à partir d'éléments tirés de $K$: c'est en ce fait que consiste la loi générale de réciprocité de Artin.

L'extension abélienne maxima d'un corps de nombres algébriques est une extension infinie; son groupe de Galois peut donc être considéré comme un groupe topologique compact,[3] qui posséde la structure topologique d'un ensemble cantorien. Plutôt que de chercher à atteindre directement ce groupe, nous avons trouvé préférable de déterminer son groupe des caractères: il est bien connu depuis les travaux de Pontrjagin qu'un groupe compact et son groupe des caractères se déterminent réciproquement.

Le résultat est qu'il existe un isomorphisme entre ce groupe des caractères et le groupe des *différentielles* du corps $K$, où nous entendons par différentielle un caractère du groupe des idèles de $K$, égal à 1 sur le groupe des idèles principaux. L'objet de la théorie du corps de classes est donc atteint puisque les différentielles sont des objets dont la connaissance ne depend que de celle du corps $K$ lui-même.

Les extensions abéliennes finies correspondent aux sous-groupes d'indices finis du groupe de Galois de l'extension abélienne maxima, c'est-à-dire encore à ceux des sous-groupes d'indices finis du groupe des idèles qui contiennent le groupe des idèles principaux. En repassant au langage de la théorie des idéaux, on retrouve la loi générale de réciprocité dans la forme qu'Artin lui a donnée.

Les §1 et 2 introduisent respectivement le groupe de Galois de l'extension abélienne maxima et le groupe des idèles. Le §3 contient la preuve d'une généralisation du théorème de Dirichlet sur les unités d'un corps; l'idée de cette généralisation est dûe à Monsieur Hasse.[4] (On montre ici qu'on peut accéder directement à cette généralisation sans passer par le théorème de Dirichlet lui-même). Le §4 introduit la notion de différentielle. Le §5 peut donner une idée d'ensemble de la démonstration du théorème central qui s'y trouve énoncé. Les §§6, 7 contiennent la preuve classique de la "première inégalité fondamentale" de la théorie. Le §8 contient certains calculs: il peut être omis par le lecteur qui ne se soucie que des ideés générales et non des technicalités. Le §9 contient la preuve sans "moyens transcendants" de la "deuxième inégalité fondamentale" dans le cas des extensions kummériennes. Le reste du mémoire suit de plus ou moins près la methode déjà indiquée dans ma thèse.

## 1. Le groupe $\mathfrak{G}_K$

$K$ étant un corps parfait (par exemple un corps de nombres algébriques), nous désignerons par $A_K/K$ l'extension composée de toutes les extensions abéliennes de $K$, que nous appellerons encore *extension abélienne maxima* de $K$.

[3] Voir Krull, *Galoissche Theorie der unendlichen algebraischen Erweiterungen*, Math. Ann. 100, 1928, p. 687.

[4] Voir les notes du cours de Monsieur Hasse, semestre d'été 1933, Université de Marburg.

Nous désignerons par $\mathfrak{G}_K$ le groupe de Galois de l'extension $A_K/K$. $\mathfrak{G}_K$ est donc un groupe abélien. Nous considérerons de plus $\mathfrak{G}_K$ comme un groupe topologique,[5] dans lequel un système fondamental de voisinages de l'unité est constitué par les groupes de Galois des extensions $A_K/Z$, $Z$ parcourant les extensions abéliennes finies de $K$.

On sait que le groupe topologique ainsi obtenu est compact, et qu'il a la structure d'un ensemble parfait discontinu.[6]

Nous appellerons caractère d'un groupe topologique abélien un homomorphisme continu de ce groupe dans le groupe multiplicatif $\Gamma$ des nombres complexes de valeur absolue 1. Les caractères d'un groupe abélien forment eux-mêmes un groupe abélien.

Le groupe $\mathfrak{G}_K$ étant partout discontinu, il en résulte que tout caractère de ce groupe est d'ordre fini.[7]

$\chi$ étant un caractère de $\mathfrak{G}_K$, soit $\mathfrak{H}$ le groupe des $s \in \mathfrak{G}_K$ tels que $\chi(s) = 1$; $\mathfrak{H}$ étant un sous-groupe fermé de $\mathfrak{G}_K$, il lui correspond par la théorie de Galois un sous-corps de $A_K$ que nous désignerons toujours au cours de ce travail par $Z_\chi$. L'extension $Z_\chi/K$ est finie, cyclique, de degré égal à l'ordre de $\chi$, comme il resulte du fait que son groupe de Galois est isomorphe au groupe $\mathfrak{G}_K / \mathfrak{H}$, qui est appliqué isomorphiquement par $\chi$ sur un sous-groupe fini, donc cyclique, de $\Gamma$. Inversement, si $Z/K$ est une extension cyclique, il existe un caractère $\chi$ de $\mathfrak{G}_K$ tel que $Z = Z_\chi$.

Soit maintenant $\Omega/K$ une extension quelconque de $K$. Il est clair que $\Omega A_K \subset A_\Omega$.[8] Tout élément $\bar{s}$ de $\mathfrak{G}_\Omega$ donne un automorphisme de $A_\Omega$ qui laisse invariants les éléments de $K$. Cet automorphisme permute donc entre eux les éléments de $A_K$ et induit dans $A_K$ un automorphisme $s$. L'application $\bar{s} \to s$ est un homomorphisme (continu) de $\mathfrak{G}_\Omega$ dans $\mathfrak{G}_K$. $\chi$ étant un caractère de $\mathfrak{G}_K$, la fonction $\bar{\chi}$ sur $\mathfrak{G}_\Omega$ définie par la formule $\bar{\chi}(\bar{s}) = \chi(s)$ est un caractère de $\mathfrak{G}_\Omega$. Nous désignerons ce caractère par la notation $N_{K/\Omega}(\chi)$(norme de $K$ à $\Omega$ de $\chi$). L'application $\chi \to N_{K/\Omega}(\chi)$ est un homomorphisme du groupe des caractères de $\mathfrak{G}_K$ dans celui des caractères de $\mathfrak{G}_\Omega$. De plus, si $\Omega \subset \Omega_1$, on a $N_{K/\Omega_1}(\chi) = N_{\Omega/\Omega_1}(N_{K/\Omega}(\chi))$.

Considérons maintenant, un isomorphisme $\tau$ de $\Omega$ avec un corps $\Omega^\tau$; nous supposerons que $\tau$ laisse invariants les éléments de $K$. Il résulte d'un théorème de la théorie de Galois que $\tau$ se laisse prolonger par un isomorphisme $\bar{\tau}$ de $A_\Omega$ avec un corps, qui est évidemment l'extension abélienne maxima $A_{\Omega^\tau}$ de $\Omega^\tau$. Considérons maintenant un élément $s$ quelconque de $\mathfrak{G}_\Omega$ ; l'application $\bar{\tau}s\bar{\tau}^{-1}$

[5] Voir Krull, *Galoissche Theorie der unendlichen algebraischen Erweiterungen*, Math. Ann. 100, 1928, p. 687.

[6] Ce resultat est prouvé par Krull dans le cas, qui est celui qui nous interesse, où $A_K$ peut être obtenu par la composition d'une infinité dénombrable d'extensions finies. Dans le cas général, on voit aisément que le groupe $\mathfrak{G}_K$ est compact, à condition de donner au mot *compact* l'acception de *bicompact*.

[7] Voir le livre de Pontrjagin, *Topological groups*, (Princeton series).

[8] Nous supposerons toujours les extensions que nous considérerons contenues dans une même extension algébriquement fermée.

est une application de $A_{\Omega^\tau}$ dans lui-même, qui est évidemment un automorphisme du corps $A_{\Omega^\tau}$ et qui laisse invariants les éléments de $\Omega^\tau$. Nous désignerons cette application par $s^{\bar\tau}$. La correspondace $s \to s^{\bar\tau}$ est un isomorphisme de $\mathfrak{G}_\Omega$ avec $\mathfrak{G}_{\Omega^\tau}$. Or je dis que $s^{\bar\tau}$ ne dépend que de $\tau$; en effet, si $\bar\tau'$ est un autre isomorphisme de $A_\Omega$ avec $A_{\Omega^\tau}$ prolongeant $\tau$, on a $\bar\tau' = \bar\tau t$ où $t$ est un élément de $\mathfrak{G}_\Omega$ ; d'où $s^{\bar\tau'} = \bar\tau t s t^{-1}\bar\tau^{-1} = \bar\tau s \bar\tau^{-1} = s^{\bar\tau}$ puisque $\mathfrak{G}_\Omega$ est abélien. Nous pouvons donc désigner $s^{\bar\tau}$ par la notation simplifié $s^\tau$.

L'isomorphisme $s \to s^\tau$ fait correspondre à chaque caractère $\chi_\Omega$ de $\mathfrak{G}_\Omega$ un caractère $\chi^\tau_{\Omega^\tau}$ de $\mathfrak{G}_{\Omega^\tau}$, défini par la formule $\chi^\tau_{\Omega^\tau}(s^\tau) = \chi_\Omega(s)$. La correspondance $\chi_\Omega \to \chi^\tau_{\Omega^\tau}$ est un isomorphisme du groupe des caractères de $\mathfrak{G}_\Omega$ avec celui des caractères de $\mathfrak{G}_{\Omega^\tau}$.

Supposons maintenant que $\Omega/K$ soit une extension cyclique. Pour tout élément $\tau$ du groupe de Galois de $\Omega/K$, on a $A_{\Omega^\tau} = A_\Omega$, et l'extension $A_\Omega/K$ est galoisienne. Soit $\mathfrak{H}$ son groupe de Galois; soient $\sigma$ un générateur du groupe de Galois de $\Omega/K$ et $\bar\sigma$ un élément de $\mathfrak{H}$ induisant dans $\Omega$ l'automorphisme $\sigma$. On a $\mathfrak{G}_\Omega \subset \mathfrak{H}$, et $\mathfrak{H}$ est engendré par $\mathfrak{G}_\Omega$ et par $\bar\sigma$. Soit $\mathfrak{C}$ le sous-groupe de $\mathfrak{H}$ engendré par les éléments $s^\tau s^{-1} = \bar\tau s \bar\tau^{-1} s^{-1}$ ($\tau$ appartenant au groupe de Galois de $\Omega/K$). Les commutateurs des éléments de $\mathfrak{G}_\Omega$ entre eux étant égaux à 1, $\mathfrak{C}$ est un sous-groupe invariant engendré par les commutateurs formés à partir d'un système de générateurs de $\mathfrak{H}$, donc $\mathfrak{C}$ est le groupe des commutateurs de $\mathfrak{H}$. Soit $A'$ le corps des éléments laissés invariants par les éléments de $\mathfrak{C}$. Le groupe de Galois de $A_\Omega/A'$ est l'adhérence[9] $\bar{\mathfrak{C}}$ de $\mathfrak{C}$; celui de $A'/K$ est $\mathfrak{H}/\bar{\mathfrak{C}}$, qui est abélien. Donc $A' \subset A_K$. Inversement, le groupe de Galois de $A_\Omega/A_K$, ayant dans $\mathfrak{H}$ un groupe quotient abélien, contient $\mathfrak{C}$, d'où $A_K \subset A'$, et $A_K = A'$. De là résulte le théorème suivant:

THEOREME 1. *Si $\Omega/K$ est une extension cyclique, une condition nécessaire et suffisante pour qu'un caractère $\chi_\Omega$ de $\mathfrak{G}_\Omega$ puisse se mettre sous la forme $N_{K/\Omega}(\chi)$, $\chi$ étant un caractère de $\mathfrak{G}_K$, est que $\chi_\Omega = \chi^\tau_\Omega$ pour tout élément $\tau$ du groupe de Galois de $\Omega/K$.*[10]

En effet, pour que $\chi_\Omega$ puisse se mettre sous la forme $N_{K/\Omega}(\chi)$, il faut et suffit que $Z_{\chi_\Omega}$ soit contenu dans $A_K$.

## 2. Le groupe fondamental

Soit $K$ un corps de nombres algébriques de degré fini. Pour chaque diviseur premier[11] $\mathfrak{p}$ de $K$ (fini ou infini), nous avons un corps de nombres $\mathfrak{p}$-adiques $K_\mathfrak{p}$. Nous désignerons par $K^*_\mathfrak{p}$ le groupe multiplicatif des nombres $\neq 0$ de $K_\mathfrak{p}$. Soit $G$ le produit direct de tous les groupes $K^*_\mathfrak{p}$. Chaque élément $\mathfrak{a}$ de $G$ est donc determiné par ses coordonnées $\mathfrak{a}_\mathfrak{p}$, où $\mathfrak{a}_\mathfrak{p} \in K^*_\mathfrak{p}$.

Nous introduirons le sous-groupe de $G$ composé des éléments $\mathfrak{a}$ tels, pour

[9] Nous entendons par là le plus petit ensemble fermé contenant $\mathfrak{C}$.

[10] Le théorème ne serait plus vrai en général pour une extension $\Omega/K$ non cyclique.

[11] Les diviseurs premiers finis sont les idéaux premiers. Pour la notion de diviseur premier infini, voir *Thèse*, Chap. III, p. 389.

presque tous les $\mathfrak{p}$, $\mathfrak{a}_\mathfrak{p}$ soit une unité $\mathfrak{p}$-adique. C'est ce groupe que nous appellerons *groupe fondamental* de $K$. Nous le désignerons par $J_K$. Ses éléments seront appelés les *idèles* de $K$.

$\mathfrak{a}$ étant un idèle, $\mathfrak{p}$ un diviseur premier, l'idèle dont toutes les coordonnées sont égales à 1, sauf la $\mathfrak{p}$-coordonnée, qui est $\mathfrak{a}_\mathfrak{p}$, sera appelé le $\mathfrak{p}$-*composant* de $\mathfrak{a}$. Les $\mathfrak{p}$-composants de tous les idèles forment un sous-groupe de $J_K$ évidemment isomorphe à $K_\mathfrak{p}^*$. En fait, il nous sera le plus souvent commode de désigner par la même notation $\mathfrak{a}_\mathfrak{p}$ le $\mathfrak{p}$-composant et la $\mathfrak{p}$-coordonnée de $\mathfrak{a}$; de même, il nous arrivera de parler de $K_\mathfrak{p}^*$ comme d'un sous-groupe de $J_K$. Nous désignerons par $U_\mathfrak{p}$ le groupe des unités $\mathfrak{p}$-adiques (égal à $K_\mathfrak{p}^*$ si $\mathfrak{p}$ est un diviseur premier infini), et nous nous donnerons aussi le droit de considérer $U_\mathfrak{p}$ comme un sous-groupe de $J_K$.

Un idèle égal à l'un de ses composants sera dit *primaire*; si on veut préciser de quel composant il s'agit, on dira "primaire pour $\mathfrak{p}$."

Soit maintenant $\Omega/K$ une extension finie de $K$. $\mathfrak{A}$ étant un idèle de $\Omega$, nous désignerons par $N_{\Omega/K}(\mathfrak{A})$ l'idèle de $K$ dont les coordonnées sont données par la formule

$$(N_{\Omega/K}(\mathfrak{A}))_\mathfrak{p} = \prod_\mathfrak{P} N_{\Omega_\mathfrak{P}/K_\mathfrak{p}}(\mathfrak{A}_\mathfrak{P})$$

le produit au second membre étant étendu aux diviseurs premiers $\mathfrak{P}$ de $\mathfrak{p}$ dans $\Omega$. On a évidemment $N_{\Omega/K}(\mathfrak{AB}) = N_{\Omega/K}(\mathfrak{A}).\ N_{\Omega/K}(\mathfrak{B})$, de sorte que l'application $N_{\Omega/K}$ est un homomorphisme de $J_\Omega$ dans $J_K$. Si $\Omega_1/\Omega$ est une extension finie de $\Omega$, et si $\mathfrak{A}_1 \in J_{\Omega_1}$ on a $N_{\Omega_1/K}(\mathfrak{A}_1) = N_{\Omega/K}(N_{\Omega_1/\Omega}(\mathfrak{A}_1))$, comme on le voit tout de suite en supposant d'abord $\mathfrak{A}_1$ primaire.

$\alpha$ étant un nombre $\neq 0$ de $K$, il n'y a qu'un nombre fini de $\mathfrak{p}$ pour lesquels $\alpha$ n'est pas une unité de $K_\mathfrak{p}$. Il y a donc un idèle dont toutes les coordonnées sont égales à $\alpha$. Si $\mathfrak{a}_\alpha$ est cet idèle, la correspondance $\alpha \to \mathfrak{a}_\alpha$ est un isomorphisme du groupe multiplicatif des nombres $\neq 0$ de $K$ avec un sous-groupe de $J_K$, que nous appellerons *sous-groupe principal*, et que nous désignerons par $P_K$. Les éléments de $P_K$ seront appelés *idèles principaux*. Nous allons montrer que, si $\Omega$ est une extension finie de $K$ et $\mathsf{A}$ un nombre de $\Omega$, on a $\mathfrak{a}_{N_{\Omega/K}\mathsf{A}} = N_{\Omega/K}(\mathfrak{A}_\mathsf{A})$, de sorte que notre définition de la norme des idèles est en accord avec celle de la norme des nombres. Il suffit de montrer que $(N_{\Omega/K}(\mathsf{A}))_\mathfrak{p} = \prod_\mathfrak{P} N_{\Omega_\mathfrak{P}/K_\mathfrak{p}}(\mathsf{A})$. Or, considérons $\Omega$ comme une algèbre sur $K$; on sait que l'algebre $\Omega_{K_\mathfrak{p}}$ qu'on en déduit en étendant à $K_\mathfrak{p}$ le corps des coefficients est isomorphe à la somme directe des $\Omega_\mathfrak{P}$,[12] le nombre $\mathsf{A}$ correspondant dans cette isomorphisme à l'élément de la somme directe des $\Omega_\mathfrak{P}$ dont les coordonnées sont égales à $\mathsf{A}$; la formule en résulte immédiatement.

En fait, il y aura intérêt dans la suite à ne pas distinguer entre les nombres $\neq 0$ de $K$ et les idèles principaux correspondants.

Si $\tau$ est un isomorphisme entre $K$ et l'un de ses conjugués $K^\tau$, $\tau$ fait corres-

[12] Soit $f(x) = 0$ l'équation irréductible dans $K$ à laquelle satisfait $\mathsf{A}$. Le polynôme $f(x)$ se décompose dans $K_\mathfrak{p}$ en le produit $\Pi_\mathfrak{P} f_\mathfrak{p}(x)$, où $f_\mathfrak{p}(x) = 0$ est l'équation irréductible à laquelle satisfait dans $K_\mathfrak{p}$ le nombre $\mathsf{A}$, considéré comme élément de $\Omega_\mathfrak{P}$.

pondre à chaque diviseur premier $\mathfrak{p}$ de $K$ un diviseur premier $\mathfrak{p}^\tau$ de $K^\tau$, et se prolonge par un isomorphisme (continu) entre $K_\mathfrak{p}$ et $K^\tau_{\mathfrak{p}^\tau}$, isomorphisme que nous désignerons encore par $\tau$. $\mathfrak{a}$ étant un idèle de $K$, nous désignerons par $\mathfrak{a}^\tau$ l'idèle de $K^\tau$ dont les coordonnées sont données par la formule $(\mathfrak{a}^\tau)_{\mathfrak{p}^\tau} = (\mathfrak{a}_\mathfrak{p})^\tau$; la correspondance $\mathfrak{a} \to \mathfrak{a}^\tau$ est un isomorphisme entre $J_K$ et $J_{K^\tau}$.

*Topologie dans* $J_K$. Soient $E$ un ensemble fini de diviseurs premiers de $K$, et $n$ un entier $> 0$. Nous désignerons par $J_K^{E,n}$ le groupe des idèles $\mathfrak{a}$ de $K$ tels que: 1) si $\mathfrak{p} \notin E$, $\mathfrak{a}_\mathfrak{p}$ soit une unité de $K_\mathfrak{p}$ ; 2) si $\mathfrak{p} \in E$, $\mathfrak{a}_\mathfrak{p}$ soit une puissance n-ième éxacte dans $K_\mathfrak{p}$. Nous poserons $J_K^E = J_K^{E,1}$, $P_K^{E,n} = P_K \cap J_K^{E,n}$ et $P_K^E = P_K \cap J_K^E$.

Nous désignerons par $G_K^{E,n}$ le groupe $J_K^n J_K^{E,n}$, où $J_K^n$ est le groupe des n-ièmes puissances des éléments de $J_K$. Si $E'$ est un autre ensemble fini de diviseurs premiers, et $n'$ un autre entier $> 0$, on a

$$G_K^{E \cup E', nn'} \subset G_K^{E,n} \cap G_K^{E',n'}$$

de sorte que l'on peut définir dans $J_K$ une structure de groupe topologique en convenant que les $G_K^{E,n}$ (pour tous les choix de $E$, $n$) constituent un système fondamental de voisinages de l'unité.[13] Nous considérerons toujours $J_K$ comme un groupe topologique, muni de la topologie qui vient d'y être définie.

Cette structure topologique n'est pas *séparée*.[14] En effet, le groupe $L = \bigcap_{E,n} G_K^{E,n}$ contient le groupe $L'$ des idèles $\mathfrak{a}$ satisfaisant aux conditions-suivantes; 1) si $\mathfrak{p}$ est fini, $\mathfrak{a}_\mathfrak{p} = 1$; 2) si $\mathfrak{p}$ est infini réel, $\mathfrak{a}_\mathfrak{p}$ est positif. Il est d'ailleurs facile de voir que $L = L'$; en effet soit $\mathfrak{a} \in L$; si $\mathfrak{p}$ est fini, $\mathfrak{a}_\mathfrak{p}$ est une puissance n-ième quelque soit $n$, d'où résulte d'abord que $\mathfrak{a}_\mathfrak{p}$ est une unité; de plus, pour tout entier $a > 0$, le groupe des classes (mod $\mathfrak{p}^a$) d'unités $\mathfrak{p}$-adiques est d'ordre fini $N_a$. $\mathfrak{a}_\mathfrak{p}$ étant puissance $N_a$-ième exacte, on a $\mathfrak{a}_\mathfrak{p} \equiv 1 \pmod{\mathfrak{p}^a}$ quel que soit $a$, d'où $\mathfrak{a}_\mathfrak{p} = 1$. Si $\mathfrak{p}$ est infini réel, $\mathfrak{a}_\mathfrak{p}$, étant un carré, est positif. D'où $\mathfrak{a} \in L'$, et $L \subset L'$.

$\mathfrak{p}$ étant un diviseur premier fini, la structure topologique induite par $J_K$ sur $K_\mathfrak{p}^*$ comporte comme système fondamental de voisinages de l'unité le système des $(K_\mathfrak{p}^*)^n$, $(K_\mathfrak{p}^*)^n$ désignant le groupe des $n$-ièmes puissances des éléments de $K_\mathfrak{p}^*$. Elle *ne coïncide pas* avec la topologie induite dans $K_\mathfrak{p}^*$ par la topologie $\mathfrak{p}$-adique de $K_\mathfrak{p}$ (elle est moins fine que cette dernière). Toutes les fois que nous parlerons de $K_\mathfrak{p}^*$ comme d'un espace topologique, c'est de la topologie induite par $J_K$ qu'il s'agira.[15]

Soit $E_i (1 \leqq i < \infty)$ une suite croissante d'ensembles finis de diviseurs pre-

[13] Pour qu'une famille de sous-groupes puisse être considérée comme un système fondamental de voisinages de l'unité d'une topologie d'un groupe abélien, il suffit que l'intersection de deux groupes de la famille en contienne toujours un troisième.

[14] Nous entendons par là que l'axiome de séparation de Hausdorf n'y est pas vérifié.

[15] La topologie que nous introduisons ici est adaptée aux questions qui font intervenir la structure multiplicative, tandisque la topologie ordinaire est plus adaptée aux questions de nature additive.

miers de $K$, telle que tout diviseur premier appartienne à l'un au moins des $E_i$. On a $\mathfrak{a} = \lim_{i\to\infty} \prod_{\mathfrak{p}\in E_i} \mathfrak{a}_\mathfrak{p}$, ce qui nous permettra d'écrire

$$\mathfrak{a} = \prod_\mathfrak{p} \mathfrak{a}_\mathfrak{p}$$

le produit étant étendu à tous les diviseurs premiers $\mathfrak{p}$.

$\Omega/K$ étant une extension finie de $K$, l'application $N_{\Omega/K}$ de $J_\Omega$ dans $J_K$ est évidemment un homomorphisme *continu* de $J_\Omega$ dans $J_K$. $\tau$ étant un isomorphisme de $\Omega$ avec l'un de ses conjugués, l'application $\mathfrak{a} \to \mathfrak{a}^\tau$ est un isomorphisme *continu* de $J_\Omega$ avec $J_{\Omega^\tau}$.

### 3. La structure du groupe $P_K^E$

$\mathfrak{p}$ étant un diviseur premier fini de $K$, nous désignerons par $w_\mathfrak{p}$ celle des valuations $\mathfrak{p}$-adiques de $K$ qui est définie par la formule $w_\mathfrak{p}(\alpha) = (N\mathfrak{p})^{-x}$, où $N\mathfrak{p}$ est la norme absolue de $\mathfrak{p}$ et $x$ l'ordre en $\mathfrak{p}$ de $\alpha$. Si $\mathfrak{p}$ est un diviseur premier infini réel, il lui correspond un isomorphisme $I_\mathfrak{p}$ de $K$ dans le corps des nombres réels; nous poserons alors $w_\mathfrak{p}(\alpha) = |I_\mathfrak{p}(\alpha)|$; enfin, si $\mathfrak{p}$ est un diviseur premier infini imaginaire, il lui correspond deux isomorphismes $I_\mathfrak{p}$, $\bar{I}_\mathfrak{p}$ de $K$ dans le corps des nombres complexes; nous poserons dans ce cas $w_\mathfrak{p}(\alpha) = |I_\mathfrak{p}(\alpha)|^2 = |\bar{I}_\mathfrak{p}(\alpha)|^2$. On a donc

$$|N(\alpha)| = \prod_{\mathfrak{p}\ \text{fini}} w_\mathfrak{p}^{-1}(\alpha) = \prod_{\mathfrak{p}\ \text{infini}} w_\mathfrak{p}(\alpha)$$

THÉORÈME 2. *Faisons correspondre à chaque diviseur premier $\mathfrak{p}$ un nombre $a_\mathfrak{p} > 0$ de telle manière que*: 1) *pour presque tous les $\mathfrak{p}$ on ait $a_\mathfrak{p} = 1$*; 2) *si $\mathfrak{p}$ est fini, $a_\mathfrak{p}$ soit une valeur prise par $w_\mathfrak{p}$ dans $K$*; 3) *on ait $\prod_\mathfrak{p} a_\mathfrak{p} \geqq |\Delta|^{1/2}$, où $\Delta$ désigne le discriminant du corps $K$. Dans ces conditions, il existe un nombre $\alpha \neq 0$ de $K$ satisfaisant à toutes les inégalités $w_\mathfrak{p}(\alpha) \leqq a_\mathfrak{p}$.*

Soit $\mathfrak{M}$ l'ensemble des nombres $\beta \in K$ satisfaisant à toutes les inégalités $w_\mathfrak{p}(\beta) \leqq a_\mathfrak{p}$ pour tous les $\mathfrak{p}$ *finis*. On sait que $\mathfrak{M}$ est un module[16] qui possède par rapport à l'anneau des entiers rationels une base $\{\theta_1, \theta_2, \cdots, \theta_n\}$ (où $n$ est le degré absolu de $K$). Les nombres $\alpha$ sont ceux qui peuvent se mettre sous la forme $\sum_{i=1}^n x_i \theta_i$, les $x_i$ étant des entiers non tous nuls tels que soient vérifiées les inégalités $\left|\sum_{i=1}^n x_i I_\mathfrak{p}(\theta_i)\right| \leqq a_\mathfrak{p}$ pour tous les $\mathfrak{p}$ *infinis*. Or, soit $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ une base des entiers de $K$; si on a $\theta_i = \sum_{j=1}^n c_{i,j}\lambda_j$ $(1 \leqq i \leqq n)$, on sait que le déterminant $|c_{i,j}|$ est égal en valeur absolue à $\prod_{\mathfrak{p}\ \text{fini}} a_\mathfrak{p}^{-1}$; il en résulte que la valeur absolue du déterminant $D$ des $n$ formes lineaires $\left|\sum_{i=1}^n x_i I_\mathfrak{p}(\theta_i)\right|$, $\left|\sum_{i=1}^n x_i \bar{I}_\mathfrak{p}(\theta_i)\right|$ est égale à $|\Delta|^{1/2} \prod_{\mathfrak{p}\ \text{fini}} a_\mathfrak{p}^{-1}$. D'où $\prod_{\mathfrak{p}\ \text{inf}} a_\mathfrak{p} \geqq |D|$; il

[16] C'est un idéal entier ou fractionnaire.

résulte alors du lemme de Minkowski qu'il existe un système d'entiers $x_i$ non tous nuls répondant à la question, ce qui démontre le théorème.

REMARQUE. Il n'existe évidemment qu'un nombre fini de systèmes d'entiers donnant lieu à nos inégalités, de sorte qu'il n'existe qu'un nombre fini de nombres $\alpha$ satisfaisant aux conditions du théorème.

Soit maintenant $E$ un ensemble fini de diviseurs premiers contenant tous les diviseurs premiers infinis. Soit $\mu$ un nombre superieur aux normes absolues des diviseurs premiers finis de $E$ et à 1. Les idèles $\mathfrak{b}$ qui satisfont simultanément à toutes les conditions

$$(1) \qquad w_\mathfrak{p}(\mathfrak{b}) \leqq 1 \text{ si } \mathfrak{p} \notin E; \qquad \prod_{\mathfrak{p} \notin E} w_\mathfrak{p}(\mathfrak{b}) \geqq 1/\mu \,|\Delta|^{-1/2}$$

se répartissent évidemment en un nombre fini de classes (mod $J_K^E$). Nous choisirons un système de représentants de ces classes, en nous arrangeant pour que, toutes les fois que l'une de ces classes contient un élément de $P_K$, le représentant choisi soit dans $P_K$. Soient $\beta_1, \beta_2, \cdots, \beta_l, \mathfrak{b}_1, \mathfrak{b}_2, \cdots, \mathfrak{b}_{l'}$ les représentants choisis, les $\beta_i$ étant dans $P_K$.

Nous choisirons d'autre part dans $E$ un diviseur premier quelconque $\mathfrak{p}_0$.

1) Soit $\mathfrak{a}$ un idèle quelconque. On peut, en vertu du théorème 1, trouver un nombre $\alpha \neq 0$ de $K$ satisfaisant aux conditions suivantes: $w_\mathfrak{p}(\alpha) \leqq w_\mathfrak{p}(\mathfrak{a})$ si $\mathfrak{p} \notin E$; $w_\mathfrak{p}(\alpha) \leqq 1$ si $\mathfrak{p} \in E$, $\mathfrak{p} \neq \mathfrak{p}_0$; $w_{\mathfrak{p}_0}(\alpha) \leqq \mu \,|\Delta|^{1/2} \prod_{\mathfrak{p} \notin E} w_\mathfrak{p}^{-1}(\mathfrak{a})$. Si $\mathfrak{b} = \alpha\mathfrak{a}^{-1}$, $\mathfrak{b}$ satisfait aux conditions (1), d'où résultent les corollaires suivants:

COROLLAIRE 1. *Si $E$ est un ensemble fini de diviseurs premiers contenant les diviseurs premiers infinis, $P_K J_K^E$ est d'indice fini dans $J_K$.*

COROLLAIRE 2. *Il existe un ensemble fini $F$ de diviseurs premiers tel que* $P_K J_K^F = J_K$.

2) Pour chaque diviseur premier $\mathfrak{p}$ de $E$ autre que $\mathfrak{p}_0$, déterminons un nombre $b_\mathfrak{p}$ tel que $0 < b_\mathfrak{p} < \min_{1 \leqq i \leqq l} w_\mathfrak{p}(\beta_i)$, et que de plus, si $\mathfrak{p}$ est fini, $b_\mathfrak{p}$ soit une valeur prise par la fonction $w_\mathfrak{p}$ dans $K$; posons $b_{\mathfrak{p}_0} = \mu \,|\Delta|^{1/2} \prod_{\substack{\mathfrak{p} \neq \mathfrak{p}_0 \\ \mathfrak{p} \in E}} b_\mathfrak{p}^{-1}$. En vertu du théorème 1, il existe un nombre $\beta \neq 0$ de $K$ tel que $w_\mathfrak{p}(\beta) \leqq b_\mathfrak{p}$ pour $\mathfrak{p} \in E$ et $w_\mathfrak{p}(\beta) \leqq 1$ pour $\mathfrak{p} \notin E$. On a $\prod_{\mathfrak{p} \notin E} w_\mathfrak{p}(\beta) = \prod_{\mathfrak{p} \in E} w_\mathfrak{p}^{-1}(\beta) \geqq \prod_{\mathfrak{p} \in E} b_\mathfrak{p}^{-1} = 1/\mu \,|\Delta|^{-1/2}$; il en résulte que $\beta$ appartient (mod $J_K^E$) à la même classe que l'un des $\beta_i$; il existe donc un indice $i$ tel que $\beta\beta_i^{-1} = \epsilon$ soit dans $P_K^E$. Il existe donc dans $P_K^E$ un nombre $\epsilon$ tel que l'on ait $w_\mathfrak{p}(\epsilon) < 1$ pour $\mathfrak{p} \neq \mathfrak{p}_0$, $\mathfrak{p} \in E$.

Désignons maintenant par $\mathfrak{p}_0, \mathfrak{p}_1, \cdots, \mathfrak{p}_h$ les diviseurs premiers de $E$. Posons $w_k = w_{\mathfrak{p}_k} (0 \leqq k \leqq h)$, et déterminons pour chaque $k$ un nombre $\epsilon_k$ de $P_K^E$ tel que $w_j(\epsilon_k) < 1$ pour $j \neq k$. Nous allons montrer que *les nombres* $\epsilon_1, \epsilon_2, \cdots, \epsilon_h$ *ne sont liés par aucune relation.* Supposons en effet que $\prod_{k=0}^{h} \epsilon_k^{u_k} = 1$, les $u_k$ étant des entiers non tous nuls, avec $u_0 = 0$. Posons $w_+ = \prod_+ w_k$, $w_- = \prod_- w_k$, l'indice $+$ signifiant que le produit est étendu aux indices $k$ tels que $u_k \geqq 0$, et l'indice $-$ qu'il est étendu aux indices $k$ tels que $u_k < 0$. On aurait

miers de $K$, telle que tout diviseur premier appartienne à l'un au moins des $E_i$. On a $\mathfrak{a} = \lim_{i\to\infty} \prod_{\mathfrak{p}\in E_i} \mathfrak{a}_\mathfrak{p}$, ce qui nous permettra d'écrire

$$\mathfrak{a} = \prod_\mathfrak{p} \mathfrak{a}_\mathfrak{p}$$

le produit étant étendu à tous les diviseurs premiers $\mathfrak{p}$.

$\Omega/K$ étant une extension finie de $K$, l'application $N_{\Omega/K}$ de $J_\Omega$ dans $J_K$ est évidemment un homomorphisme *continu* de $J_\Omega$ dans $J_K$. $\tau$ étant un isomorphisme de $\Omega$ avec l'un de ses conjugués, l'application $\mathfrak{a} \to \mathfrak{a}^\tau$ est un isomorphisme *continu* de $J_\Omega$ avec $J_{\Omega^\tau}$.

### 3. La structure du groupe $P_K^E$

$\mathfrak{p}$ étant un diviseur premier fini de $K$, nous désignerons par $w_\mathfrak{p}$ celle des valuations $\mathfrak{p}$-adiques de $K$ qui est définie par la formule $w_\mathfrak{p}(\alpha) = (N\mathfrak{p})^{-x}$, où $N\mathfrak{p}$ est la norme absolue de $\mathfrak{p}$ et $x$ l'ordre en $\mathfrak{p}$ de $\alpha$. Si $\mathfrak{p}$ est un diviseur premier infini réel, il lui correspond un isomorphisme $I_\mathfrak{p}$ de $K$ dans le corps des nombres réels; nous poserons alors $w_\mathfrak{p}(\alpha) = |I_\mathfrak{p}(\alpha)|$; enfin, si $\mathfrak{p}$ est un diviseur premier infini imaginaire, il lui correspond deux isomorphismes $I_\mathfrak{p}$, $\bar{I}_\mathfrak{p}$ de $K$ dans le corps des nombres complexes; nous poserons dans ce cas $w_\mathfrak{p}(\alpha) = |I_\mathfrak{p}(\alpha)|^2 = |\bar{I}_\mathfrak{p}(\alpha)|^2$. On a donc

$$|N(\alpha)| = \prod_{\mathfrak{p}\ \text{fini}} w_\mathfrak{p}^{-1}(\alpha) = \prod_{\mathfrak{p}\ \text{infini}} w_\mathfrak{p}(\alpha)$$

THÉORÈME 2. *Faisons correspondre à chaque diviseur premier $\mathfrak{p}$ un nombre $a_\mathfrak{p} > 0$ de telle manière que*: 1) *pour presque tous les $\mathfrak{p}$ on ait $a_\mathfrak{p} = 1$*; 2) *si $\mathfrak{p}$ est fini, $a_\mathfrak{p}$ soit une valeur prise par $w_\mathfrak{p}$ dans $K$*; 3) *on ait $\prod_\mathfrak{p} a_\mathfrak{p} \geqq |\Delta|^{1/2}$, où $\Delta$ désigne le discriminant du corps $K$. Dans ces conditions, il existe un nombre $\alpha \neq 0$ de $K$ satisfaisant à toutes les inégalités $w_\mathfrak{p}(\alpha) \leqq a_\mathfrak{p}$.*

Soit $\mathfrak{M}$ l'ensemble des nombres $\beta \in K$ satisfaisant à toutes les inégalités $w_\mathfrak{p}(\beta) \leqq a_\mathfrak{p}$ pour tous les $\mathfrak{p}$ *finis*. On sait que $\mathfrak{M}$ est un module[16] qui possède par rapport à l'anneau des entiers rationels une base $\{\theta_1, \theta_2, \cdots, \theta_n\}$ (où $n$ est le degré absolu de $K$). Les nombres $\alpha$ sont ceux qui peuvent se mettre sous la forme $\sum_{i=1}^n x_i\,\theta_i$, les $x_i$ étant des entiers non tous nuls tels que soient vérifiées les inégalités $\left|\sum_{i=1}^n x_i I_\mathfrak{p}(\theta_i)\right| \leqq a_\mathfrak{p}$ pour tous les $\mathfrak{p}$ *infinis*. Or, soit $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ une base des entiers de $K$; si on a $\theta_1 = \sum_{j=1}^n c_{i,j}\lambda_j$ $(1 \leqq i \leqq n)$, on sait que le déterminant $|c_{i,j}|$ est égal en valeur absolue à $\prod_{\mathfrak{p}\ \text{fini}} a_\mathfrak{p}^{-1}$; il en résulte que la valeur absolue du déterminant $D$ des $n$ formes lineaires $\left|\sum_{i=1}^n x_i I_\mathfrak{p}(\theta_i)\right|$, $\left|\sum_{i=1}^n x_i \bar{I}_\mathfrak{p}(\theta_i)\right|$ est égale à $|\Delta|^{1/2} \prod_{\mathfrak{p}\ \text{fini}} a_\mathfrak{p}^{-1}$. D'où $\prod_{\mathfrak{p}\ \text{inf}} a_\mathfrak{p} \geqq |D|$; il

[16] C'est un idéal entier ou fractionnaire.

résulte alors du lemme de Minkowski qu'il existe un système d'entiers $x_i$ non tous nuls répondant à la question, ce qui démontre le théorème.

REMARQUE. Il n'existe évidemment qu'un nombre fini de systèmes d'entiers donnant lieu à nos inégalités, de sorte qu'il n'existe qu'un nombre fini de nombres $\alpha$ satisfaisant aux conditions du théorème.

Soit maintenant $E$ un ensemble fini de diviseurs premiers contenant tous les diviseurs premiers infinis. Soit $\mu$ un nombre superieur aux normes absolues des diviseurs premiers finis de $E$ et à 1. Les idèles $\mathfrak{b}$ qui satisfont simultanément à toutes les conditions

$$(1) \qquad w_\mathfrak{p}(\mathfrak{b}) \leqq 1 \text{ si } \mathfrak{p} \notin E; \qquad \prod_{\mathfrak{p} \notin E} w_\mathfrak{p}(\mathfrak{b}) \geqq 1/\mu \,|\, \Delta \,|^{-1/2}$$

se répartissent évidemment en un nombre fini de classes (mod $J_K^E$). Nous choisirons un système de représentants de ces classes, en nous arrangeant pour que, toutes les fois que l'une de ces classes contient un élément de $P_K$, le représentant choisi soit dans $P_K$. Soient $\beta_1, \beta_2, \cdots, \beta_l, \mathfrak{b}_1, \mathfrak{b}_2, \cdots, \mathfrak{b}_{l'}$ les représentants choisis, les $\beta_i$ étant dans $P_K$.

Nous choisirons d'autre part dans $E$ un diviseur premier quelconque $\mathfrak{p}_0$.

1) Soit $\mathfrak{a}$ un idèle quelconque. On peut, en vertu du théorème 1, trouver un nombre $\alpha \neq 0$ de $K$ satisfaisant aux conditions suivantes: $w_\mathfrak{p}(\alpha) \leqq w_\mathfrak{p}(\mathfrak{a})$ si $\mathfrak{p} \notin E$; $w_\mathfrak{p}(\alpha) \leqq 1$ si $\mathfrak{p} \in E$, $\mathfrak{p} \neq \mathfrak{p}_0$; $w_{\mathfrak{p}_0}(\alpha) \leqq \mu \,|\, \Delta \,|^{1/2} \prod_{\mathfrak{p} \notin E} w_\mathfrak{p}^{-1}(\mathfrak{a})$. Si $\mathfrak{b} = \alpha\mathfrak{a}^{-1}$, $\mathfrak{b}$ satisfait aux conditions (1), d'où résultent les corollaires suivants:

COROLLAIRE 1. *Si $E$ est un ensemble fini de diviseurs premiers contenant les diviseurs premiers infinis, $P_K J_K^E$ est d'indice fini dans $J_K$.*

COROLLAIRE 2. *Il existe un ensemble fini $F$ de diviseurs premiers tel que* $P_K J_K^F = J_K$.

2) Pour chaque diviseur premier $\mathfrak{p}$ de $E$ autre que $\mathfrak{p}_0$, déterminons un nombre $b_\mathfrak{p}$ tel que $0 < b_\mathfrak{p} < \min_{1 \leqq i \leqq l} w_\mathfrak{p}(\beta_i)$, et que de plus, si $\mathfrak{p}$ est fini, $b_\mathfrak{p}$ soit une valeur prise par la fonction $w_\mathfrak{p}$ dans $K$; posons $b_{\mathfrak{p}_0} = \mu \,|\, \Delta \,|^{1/2} \prod_{\substack{\mathfrak{p} \neq \mathfrak{p}_0 \\ \mathfrak{p} \in E}} b_\mathfrak{p}^{-1}$. En vertu du théorème 1, il existe un nombre $\beta \neq 0$ de $K$ tel que $w_\mathfrak{p}(\beta) \leqq b_\mathfrak{p}$ pour $\mathfrak{p} \in E$ et $w_\mathfrak{p}(\beta) \leqq 1$ pour $\mathfrak{p} \notin E$. On a $\prod_{\mathfrak{p} \notin E} w_\mathfrak{p}(\beta) = \prod_{\mathfrak{p} \in E} w_\mathfrak{p}^{-1}(\beta) \geqq \prod_{\mathfrak{p} \in E} b_\mathfrak{p}^{-1} = 1/\mu \,|\, \Delta \,|^{-1/2}$; il en résulte que $\beta$ appartient (mod $J_K^E$) à la même classe que l'un des $\beta_i$; il existe donc un indice $i$ tel que $\beta\beta_i^{-1} = \epsilon$ soit dans $P_K^E$. Il existe donc dans $P_K^E$ un nombre $\epsilon$ tel que l'on ait $w_\mathfrak{p}(\epsilon) < 1$ pour $\mathfrak{p} \neq \mathfrak{p}_0$, $\mathfrak{p} \in E$.

Désignons maintenant par $\mathfrak{p}_0, \mathfrak{p}_1, \cdots, \mathfrak{p}_h$ les diviseurs premiers de $E$. Posons $w_k = w_{\mathfrak{p}_k} (0 \leqq k \leqq h)$, et déterminons pour chaque $k$ un nombre $\epsilon_k$ de $P_K^E$ tel que $w_j(\epsilon_k) < 1$ pour $j \neq k$. Nous allons montrer que *les nombres* $\epsilon_1, \epsilon_2, \cdots, \epsilon_h$ *ne sont liés par aucune relation.* Supposons en effet que $\prod_{k=0}^{h} \epsilon_k^{u_k} = 1$, les $u_k$ étant des entiers non tous nuls, avec $u_0 = 0$. Posons $w_+ = \prod_+ w_k$, $w_- = \prod_- w_k$, l'indice $+$ signifiant que le produit est étendu aux indices $k$ tels que $u_k \geqq 0$, et l'indice $-$ qu'il est étendu aux indices $k$ tels que $u_k < 0$. On aurait

$\prod_{k=0}^{h} (w_-(\epsilon_k))^{u_k} = 1$, d'où aussi, puisque $\prod_{k=0}^{h} w_k(\epsilon) = 1$ pour tout $\epsilon \in P_K^E$, $\prod_+(w_-(\epsilon_k))^{u_k}$. $\prod_-(w_+(\epsilon_k))^{-u_k} = 1$, ce qui est impossible, tous les facteurs du produit étant $< 1$.

De plus, *le groupe $P'$ engendre par $\epsilon_1, \epsilon_2, \cdots, \epsilon_h$ est d'indice fini dans $P_K^E$.* Posons en effet $a_k = w_k(\epsilon_k)$. $\epsilon$ étant un nombre de $P_K^E$ satisfaisant aux conditions $w_k(\epsilon) \leqq a_k$ (par exemple $\epsilon = 1$), considérons tous les systémes d'entiers $u_1, u_2, \cdots, u_h$ tous $\geqq 0$ tels que l'on ait les inégalités $w_k(\epsilon \cdot \epsilon_1^{u_1} \cdot \epsilon_2^{u_2} \cdots \epsilon_h^{u_h}) \leqq a_k$ $(1 \leqq k \leqq h)$. Il n'y en a qu'un nombre fini, et, parmi eux, au moins un, soit $\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_h$, qui donne à $w_0(\epsilon \cdot \epsilon_1^{u_1} \cdot \epsilon_2^{u_2} \cdots \epsilon_h^{u_h})$ sa plus petite valeur. Soit $\eta = \prod_{k=1}^{h} \epsilon_k^{\bar{u}_k}$; s'il existait un $k_0$ $(1 \leqq k_0 \leqq h)$ tel que $w_{k_0}(\epsilon\eta) \leqq 1$ on aurait aussi $w_k(\epsilon\eta\epsilon_{k_0}) \leqq a_k$ $(1 \leqq k \leqq h)$ et $w_0(\epsilon\eta\epsilon_{k_0}) < w_0(\epsilon\eta)$ ce qui est impossible. On en conclut que tout nombre $\epsilon$ de $P_K^E$ satisfaisant aux conditions $w_k(\epsilon) \leqq a_k$ $(1 \leqq k \leqq h)$ est congrû (mod $P'$) à un nombre $\epsilon\eta$ pour lequel on a $1 < w_k(\epsilon\eta) \leqq a_k (1 \leqq k \leqq h)$.

En particulier, il existe un nombre $\eta_1$ de $P'$ satisfaisant aux conditions $w_k(\eta_1) > 1$ $(1 \leqq k \leqq h)$, d'où résulte que tout nombre de $P_K^E$ est congru (mod $P'$) à un nombre $\epsilon$ tel que $w_k(\epsilon) \leqq 1$ $(1 \leqq k \leqq h)$, donc aussi à un nombre satisfaisant aux conditions $1 < w_k(\epsilon\eta) \leqq a_k$ $(1 \leqq k \leqq h)$, qui entraînent $w_0(\epsilon\eta) < 1$. Mais, en vertu de la remarque qui suit le théorème 1, il n'y a qu'un nombre fini de nombres de $P_K^E$ satisfaisant à ces conditions.

Nous avons donc demontré le

THÉORÈME 3. *$E$ étant un ensemble de $h + 1$ diviseurs premiers, contenant tous les diviseurs premiers infinis, à chaque $\mathfrak{p} \in E$ on peut faire correspondre un nombre $\epsilon_\mathfrak{p}$ de $P_K^E$ satisfaisant aux conditions $w_\mathfrak{q}(\epsilon_\mathfrak{p}) < 1$ pour tous les $\mathfrak{q}$ de $E$ autres que $\mathfrak{p}$; de plus, si on prend $h$ des nombres ainsi obtenus, ils ne sont liés par aucune relation et engendrent un sous-groupe d'indice fini de $P_K^E$.*[17]

Remarque. Ce théorème reste vrai si on remplace les valuations $w_\mathfrak{p}$ introduites au début de ce § par des valuations équivalentes quelconques.

Il en résulte tout de suite, en vertu d'un théorème bien connu sur les groupes abéliens, que $P_K^E$ est le produit direct d'un groupe fini et d'un groupe abélien libre à $h$ générateurs. Or, les éléments d'ordre fini de $P_K^E$ sont les racines de l'unité de $K$. On en tire le

COROLLAIRE. *$E$ satisfaisant aux conditions du théorème 2, $n$ étant un entier $> 0$, si $K$ contient une racine primitive $n$-ième de l'unité, l'indice dans $P_K^E$ du groupe $(P_K^E)^n$ des $n$-ièmes puissances des nombres de $P_K^E$ est égal à $n^{h+1}$.*

## 4. Notion de differentielle

Soit $K$ un corps de nombres algébriques de degré fini. Commençons par remarquer que tout caractère $\psi$ du groupe topologique $J_K$ est d'ordre fini. En

[17] La demonstration que nous avons donnée est entièrement calquée sur la démonstration du théorème de Dirichlet donnée par van der Waerden: *Ein logarithmenfreier Beweis des Dirichletschen Einheitensatzes*, Hamb. Abhand., 6, 1928, p. 259.

effet, soit $U$ un voisinage de 1 dans le groupe multiplicatif des nombres complexes de valeur absolue 1 qui ne contienne aucun sous-groupe $\neq \{1\}$. En vertu de la continuité de $\psi$, il existe un ensemble fini $E$ de diviseurs premiers de $K$ et un entier $n$ tels que $\psi(G_K^{E,n}) \subset U$ d'où $\psi(G_K^{E,n}) = \{1\}$ et *a fortiori* $\psi(J_K^n) = 1$ et $\psi^n = 1$.

DÉFINITION. *Nous dirons que le caractère $\psi$ de $J_K$ est ramifié en $\mathfrak{p}$ si $\psi$ n'est pas identiquement égal à 1 sur $U_\mathfrak{p}$.*

Un caractère ne peut avoir qu'un nombre fini de diviseurs premiers de ramification. Car, si $E$ et $n$ sont tels que $\psi(G_K^{E,n}) = \{1\}$, on a $\psi = 1$ sur $U_\mathfrak{p}$ à condition que $\mathfrak{p} \notin E$.

$\mathfrak{p}$ étant un diviseur premier quelconque, $\psi$ un caractère de $J_K$, le caractère $\psi_\mathfrak{p}$ de $K_\mathfrak{p}^*$ qui coincide avec $\psi$ sur $K_\mathfrak{p}^*$ est appelé la $\mathfrak{p}$-coordonnée de $\psi$. Le caractère $\psi$ est bien determiné quand on connait toutes ses $\mathfrak{p}$-coordonnées, car de la formule $\mathfrak{a} = \prod_\mathfrak{p} \mathfrak{a}_\mathfrak{p}$ pour un idèle quelconque $\mathfrak{a}$ et de la continuité de $\psi$ résulte que l'on a $\psi(\mathfrak{a}) = \prod_\mathfrak{p} \psi_\mathfrak{p}(\mathfrak{a}_\mathfrak{p})$.

Un caractère qui est identiquement égal à 1 sur les $K_\mathfrak{q}^*$ pour $\mathfrak{q} \neq \mathfrak{p}$ est appelé un *caractère local* pour $\mathfrak{p}$. Les notions de caractère local pour $\mathfrak{p}$ et de $\mathfrak{p}$-coordonnée d'un caractère sont pratiquement équivalentes.

DÉFINITION. *On appelle différentielle de $K$ un caractère de $J_K$ identiquement égal a 1 sur $P_K$.*[18]

Les différentielles de $K$ forment donc un groupe multiplicatif. $\Omega/K$ étant une extension finie du corps $K$, $\varphi$ une différentielle de $K$, la fonction $\varphi_\Omega$ sur $J_\Omega$ définie par la formule $\varphi_\Omega(\mathfrak{A}) = \varphi(N_{\Omega/K}(\mathfrak{A}))$ est une différentielle de $\Omega$ que nous designerons par la notation $N_{K/\Omega}(\varphi)$. L'application $N_{K/\Omega}$ est un homomorphisme du groupe des différentielles de $K$ dans celui des différentielles de $\Omega$. Si $\Omega_1/\Omega$ est une extension finie de $\Omega$, on a $N_{K/\Omega_1}(\varphi) = N_{\Omega/\Omega_1}(N_{K/\Omega}(\varphi))$.

$\tau$ étant un isomorphisme de $\Omega$ avec l'un de ses conjugués, $\varphi_\Omega$ une différentielle de $\Omega$, la fonction $\varphi_\Omega^\tau$ définie sur $J_{\Omega^\tau}$ par la formule $\varphi_\Omega^\tau(\mathfrak{A}^\tau) = \varphi_\Omega(\mathfrak{A})$ est une différentielle de $\Omega^\tau$, et l'application $\varphi_\Omega \to \varphi_\Omega^\tau$ est un isomorphisme du groupe des différentielles de $\Omega$ avec celui de $\Omega^\tau$.

Nous allons maintenant montrer *qu'une différentielle est déjà bien determinée quand on connait presque toutes ses coordonnées.* Pour cela, nous avons besoin du lemme suivant:

LEMME. *$\mathfrak{a}$ étant un idèle du corps $K$, $n$ un entier $> 0$, $E$ un ensemble fini de diviseurs premiers, il existe un nombre $\alpha \in P_K$ tel que l'on ait $(\mathfrak{a}\alpha^{-1})_\mathfrak{p} \in (K_\mathfrak{p}^*)^n$ pour tous les $\mathfrak{p} \in E$.*

Soit $\mathfrak{p}$ un diviseur premier fini de $E$. La théorie de la fonction exponentielle[19] dans $K_\mathfrak{p}$ montre qu'il existe un exposant $a_\mathfrak{p}$ tel que toute unité de $K_\mathfrak{p}$ qui est $\equiv 1 \pmod{\mathfrak{p}^a}$ soit dans $(K_\mathfrak{p}^*)^n$. D'autre part, il existe un nombre $\alpha_\mathfrak{p}$ de $K$ tel

[18] Ces différentielles jouent dans le cas des corps de nombres un rôle à certains points de vue analogue à celui joué par les différentielles dans le cas des corps de fonctions algébriques. Voir notamment: A. Weil, *Zur algebraischen Theorie der algebraischen Funktionen*, Crelle, 179, 1938, p. 129.

[19] Voir *Thèse*, Chap. V, p. 416.

que $\mathfrak{a}_\mathfrak{p}\alpha_\mathfrak{p}^{-1} \equiv 1 \pmod{\mathfrak{p}^{a_\mathfrak{p}}}$. La théorie des congruences montre qu'il existe un nombre $\alpha_0 \in K$ tel que $\alpha_\mathfrak{p}\alpha_0^{-1} \equiv 1 \pmod{\mathfrak{p}^a}$ pour tous les diviseurs premiers finis de $E$. Soit d'autre part $\theta$ un nombre engendrant le corps $K$. A chaque diviseur premier infini réel $\mathfrak{p}$ de $K$ correspond un isomorphisme $I_\mathfrak{p}$ de $K$ dans le corps des nombres réels, et les $I_\mathfrak{p}(\theta)$ sont différents deux-à-deux. Soient $\theta_1, \theta_2, \cdots, \theta_r$ les nombres $I_\mathfrak{p}(\theta)$ rangés dans un ordre tel que $\theta_1 < \theta_2 < \cdots < \theta_r$. Soient $\rho_0, \rho_1, \cdots, \rho_r$ des nombres rationnels tels que $\rho_0 < \theta_1 < \rho_1 < \theta_2 < \cdots < \theta_r < \rho_r$. Posons $\varphi_i = (\theta - \rho_{i-1})(\theta - \rho_i)$. Si $I_\mathfrak{p}(\theta) = \theta_i$, on a $I_\mathfrak{p}(\varphi_i) < 0$, et $I_\mathfrak{q}(\varphi_i) > 0$ pour tous les diviseurs premiers infinis réels $\mathfrak{q} \neq \mathfrak{p}$. On peut donc trouver un produit $\varphi$ de nombres $\varphi_i$ tel que, pour tous les $\mathfrak{p}$ infinis réels, $I_\mathfrak{p}(\varphi)$ ait le même signe que $\mathfrak{a}_\mathfrak{p}$, et la même propriété appartiendra au nombre $\alpha = \alpha_0 + N\varphi$ si $N$ est un entier rationnel suffisamment grand. On peut de plus choisir $N$ de telle manière que $\alpha_0^{-1}\alpha \equiv 1 \pmod{\mathfrak{p}^{a_\mathfrak{p}}}$ pour les diviseurs premiers finis $\mathfrak{p}$ de $E$. Le nombre $\alpha$ ainsi obtenu satisfera aux conditions du théorème, puisque tout nombre positif est une puissance n-ième exacte dans le corps des nombres réels.

Ceci posé, pour montrer qu'une différentielle est bien déterminée par la donnée de presque toutes ses coordonnées, il suffit de montrer que, $\varphi$ étant une différentielle, s'il existe un ensemble fini $E$ de diviseurs premiers tel que $\varphi_\mathfrak{p}$ soit identiquement égale à 1 si $\mathfrak{p} \notin E$, $\varphi$ est identiquement égale à 1. En remplaçant au besoin $E$ par un ensemble plus grand, on peut supposer que $\varphi(G_K^{E,n}) = \{1\}$ pour un certain entier $n > 0$. Soit $\mathfrak{a}$ un idèle quelconque. Déterminons un nombre $\alpha$ de $P_K$ tel que $(\mathfrak{a}\alpha^{-1})_\mathfrak{p} \in (K_\mathfrak{p}^*)^n$ pour tous les $\mathfrak{p} \in E$. On a $\varphi(\mathfrak{a}) = \varphi(\mathfrak{a}\alpha^{-1}) = \prod_{\mathfrak{p} \in E} \varphi_\mathfrak{p}(\mathfrak{a}\alpha^{-1}) = \prod_{\mathfrak{p} \in E} \varphi(\mathfrak{a}\alpha^{-1})_\mathfrak{p}$ ce qui prouve notre assertion.

### 5. Enonce du theoreme central. Marche generale de la demonstration

Nous avons en vue dans ce travail la démonstration du théorème suivant: *$K$ étant un corps de nombres algébriques de degré fini, il existe un isomorphisme $\chi \to \varphi = \Phi_K(\chi)$ du groupe des caractères $\chi$ du groupe de Galois $\mathfrak{G}_K$ de l'extension abélienne complète de $K$ avec le groupe des différentielles $\varphi$ de $K$, cet isomorphisme jouissant des propriétés suivantes*:

I. *Si $\Omega/K$ est une extension finie de $K$, on a*

$$\Phi_\Omega(N_{K|\Omega}\chi) = N_{K|\Omega}(\Phi_K(\chi))$$

II. *Si $\tau$ est un isomorphisme de $K$ avec l'un de ses conjugués, on a*

$$\Phi_{K^\tau}(\chi^\tau) = (\Phi_K(\chi))^\tau$$

III. *Les diviseurs premiers de ramification de $\chi$ et de $\Phi_K(\chi)$ sont les mêmes.*

(Nous convenons de dire qu'un caractère $\chi$ de $\mathfrak{G}_K$ est ramifié en $\mathfrak{p}$ si $\mathfrak{p}$ est ramifié dans $Z_\chi$).

La propriété $I$ montre que, si $\varphi = \Phi_K(\chi)$, on aura $\varphi(N_{Z_\chi/K}(\mathfrak{A})) = 1$ pour tout idèle de $Z_\chi$. Nous sommes donc conduits à la définition suivante:

DÉFINITION. *Une différentielle $\varphi$ de $K$ est dite associée à l'extension abélienne finie $Z/K$ de $K$ si $\varphi(N_{Z/K}(\mathfrak{A})) = 1$ pour tout idèle $\mathfrak{A}$ du corps $Z$.*

Les différentielles associées à une extension abélienne $Z/K$ forment donc un groupe. Nous commencerons, au cours du §7, par démontrer que, si $Z/K$ est cyclique, ce groupe est d'ordre multiple de $[Z:K]$.[20]

Nous démontrerons ensuite au §9 que l'ordre du groupe des différentielles associées à une extension abélienne finie quelconque $Z/K$ est au plus égal à $[Z:K]$.[21]

Nous définirons au cours du §11 la différentielle $\Phi_K(\chi)$ associée au caractère $\chi$ de $\mathfrak{G}_K$ dans le cas particulier où $Z_\chi/K$ est une extension circulaire.

Nous étendrons ensuite au §12 la définition de $\Phi_K(\chi)$ au cas d'un caractère $\chi$ quelconque de $\mathfrak{G}_K$ en étudiant certains corps composés de $Z_\chi$ avec des extensions circulaires, généralisant en cela la méthode employée par Artin pour démontrer la loi de réciprocité, dont l'idée centrale consiste à introduire un surcorps $\Omega$ de $K$ tel que l'extension $\Omega Z_\chi/\Omega$ soit circulaire.

La fonction $\Phi_K(\chi)$ étant ainsi définie, il ne restera plus, au §13, qu'à prouver que toute différentielle $\varphi$ de $K$ peut se mettre sous la forme $\Phi_K(\chi)$, avec un $\chi$ convenable.

### 6. Étude locale d'une extension cyclique

Soit $Z/K$ une extension cyclique du corps de nombres algébriques $K$. $\mathfrak{p}$ étant un diviseur premier du corps $K$, nous nous proposons de démontrer la formule

$$[K_\mathfrak{p}^* : K_\mathfrak{p}^* \cap N_{Z/K} J_Z] = [ZK_\mathfrak{p} : K_\mathfrak{p}]$$

Remarquons d'abord que $K_\mathfrak{p}^* \cap N_{Z/K}J_Z = N_{ZK_\mathfrak{p}/K_\mathfrak{p}}(ZK_\mathfrak{p})^*$. En effet, il est évident que le second membre est un sous-groupe du premier. Inversement, soit $\mathfrak{a} = N_{Z/K}\,\mathfrak{A}$ un idèle de $N_{Z/K}J_Z$. On a $\mathfrak{a}_\mathfrak{p} = \prod_\mathfrak{P} N_{Z_\mathfrak{P}/K_\mathfrak{p}}(\mathfrak{A}_\mathfrak{P})$, le produit étant étendu aux diviseurs premiers $\mathfrak{P}$ de $\mathfrak{p}$ dans $Z$. Les extensions $Z_\mathfrak{P}/K_\mathfrak{p}$ et $ZK_\mathfrak{p}/K_\mathfrak{p}$ étant isomorphes, chacun des facteurs $N_{Z_\mathfrak{P}/K_\mathfrak{p}}(\mathfrak{A}_\mathfrak{P})$ appartient au groupe $N_{ZK_\mathfrak{p}/K_\mathfrak{p}}(ZK_\mathfrak{p})^*$, et il en est de même de $\mathfrak{a}_\mathfrak{p}$, ce qui achève la preuve de notre formule.

Changeant légèrement les notations, il nous suffira donc de démontrer la formule $[K^*:N_{Z/K}Z^*] = [Z:K]$ pour une extension cyclique $Z/K$ d'un corps local $K$. Nous désignerons par $\mathfrak{p}$ le diviseur premier de $K$, par $\mathfrak{P}$ celui de $Z$.

Supposons d'abord $\mathfrak{p}$ fini. On a $[K^*:N_{Z/K}Z^*] = [K^*:U_K N_{Z/K}Z^*]\,[U_K N_{Z/K}Z^*:N_{Z/K}Z^*]$, où $U_K$ représente le groupe des unités de $K$. Si $f$ est le degré relatif de $\mathfrak{P}$ par rapport à $K$, les ordres pour $\mathfrak{p}$ des éléments de $N_{Z/K}Z^*$ sont les multiples de $f$, de sorte que le premier facteur est égal à $f$.

Le second facteur s'écrit $[U_K:N_{Z/K}U_Z]$, où $U_Z$ est le groupe des unités de $Z$. Soit $\sigma$ un générateur du groupe de Galois $\mathfrak{g}$ de $Z/K$. Considérons les opérations $\mathsf{E} \to N_{Z/K}\mathsf{E}$, $\mathsf{E} \to \mathsf{E}^{1-\sigma}$ comme des homomorphismes de $U_Z$ dans lui-même, que nous désignerons par $L_1$ et $L_2$. On a, pour tout $\mathsf{E} \in U_Z$, $L_1L_2\mathsf{E} =$

[20] Ce fait constitue la *première inégalité fondamentale* de la théorie du corps de classes.

[21] C'est la *seconde inégalité fondamentale* de la théorie.

$L_2L_1\mathfrak{E} = 1$. L'indice à calculer se met sous la forme $[\overset{-1}{L_2}(1):L_1(U_Z)]$. D'autre part, le lemme de Herbrand[22] nous apprend que le quotient $[\overset{-1}{L_2}(1):L_1(U_Z)].$ $[\overset{-1}{L_1}(1):L_2(U_Z)]^{-1}$ ne change pas si on remplace $U_Z$ par un sous-groupe $H$ d'indice fini dans $U_Z$ tel que $L_1H \subset H$, $L_2H \subset H$ (à condition de remplacer $\overset{-1}{L_1}(1)$ et $\overset{-1}{L_2}(1)$ par les intersections de ces groupes avec $H$).

Soit $\mathfrak{o}$ l'anneau des entiers de $K$. Nous savons qu'il existe dans $Z$ un module $\mathfrak{M}$ par rapport à $\mathfrak{o}$ qui possède une base minima par rapport à $\mathfrak{o}$ formée d'un nombre $\Gamma$ et de ses conjugués relatifs par rapport à $K$.[23] En multipliant au besoin $\Gamma$ par un nombre de $\mathfrak{o}$, on peut supposer de plus que la fonction exponentielle[24] $\mathfrak{p}$-adique est définie sur $\mathfrak{M}$, et fournit un isomorphisme du groupe additif $\mathfrak{M}$ avec un sous-groupe $H$ de $U_Z$. On peut choisir un exposant entier $N$ tel que $(U_Z)^N$ soit contenu dans le champ de définition de la fonction logarithmique[24] $\mathfrak{p}$-adique, et que $\log (U_Z)^N \subset \mathfrak{M}$, d'où $U_Z^N \subset H$. Il en résulte que $H$ est un sous-groupe d'indice fini de $U_Z$ (car on sait que $U_Z^N$ est d'indice fini dans $U_Z$ ; nous aurons d'ailleurs à calculer explicitement cet indice au §8). D'autre part, si $X \in \mathfrak{M}$, on a aussi $\mathrm{Sp}_{Z/K}X \in \mathfrak{M}$[25] et $X - \sigma X \in \mathfrak{M}$; en vertu des formules $\exp.(\mathrm{Sp}_{Z/K}X) = N_{Z/K}(\exp X)$ et $\exp(X - \sigma X) = (\exp X)^{1-\sigma}$, il résulte de là que $L_1H \subset H$, $L_2H \subset H$. Nous pouvons donc appliquer le lemme de Herbrand; tenant compte de l'isomorphisme établi par la fonction exponentielle entre $\mathfrak{M}$ et $H$, on voit que $[\overset{-1}{L_2}(1):L_1(U_Z)] \cdot [\overset{-1}{L_1}(1):L_2(U_Z)]^{-1} = [\overset{-1}{M_2}(0):M_1(\mathfrak{M})][\overset{-1}{M_1}(0):M_2(\mathfrak{M})]^{-1}$, où $M_1$ et $M_2$ sont les fonctions définies dans $\mathfrak{M}$ par les formules $M_1(X) = \mathrm{Sp}_{Z/K}X$, $M_2(X) = X - \sigma X$.

Or, le groupe $\overset{-1}{M_2}(0)$ est égal à $\mathfrak{M} \cap K$; un élément $X = \sum_{\tau \in \mathfrak{g}} u_\tau \Gamma^\tau$ $(u_\tau \in \mathfrak{o})$ ne peut être dans $K$ que si tous les $u_\tau$ sont égaux à un même élément $u$, et $X$ est alors la trace de $u\Gamma$ par rapport à $K$. On a donc $[\overset{-1}{M_2}(0):M_1(\mathfrak{M})] = 1$. Si l'élément $X = \sum_{\tau \in \mathfrak{g}} u_\tau \Gamma^\tau$ appartient à $\overset{-1}{M_1}(0)$, on a $\sum_{\tau \in \mathfrak{g}} u_\tau = 0$, de sorte qu'il existe des nombres $v_\tau \in \sigma$ tels que $u_\tau = v_{\sigma^{-1}\tau} - v_\tau$, et on a $X = Y - \sigma Y$, en posant $Y = \sum_{\tau \in \mathfrak{g}} v_\tau \Gamma^\tau$. Il en résulte l'égalité $[\overset{-1}{M_1}(0):M_2(\mathfrak{M})] = 1$.

On en conclut que $[U_K:N_{Z/K}(U_Z)] = [\overset{-1}{L_1}(1):L_2(U_Z)]$. Or, en vertu du théorème normique de Hilbert, on a $\overset{-1}{L_1}(1) = (Z^*)^{1-\sigma}$, de sorte que notre indice s'écrit $[(Z^*)^{1-\sigma}:(U_Z)^{1-\sigma}] = [Z^*:K^*U_Z]$. Mais, si $e$ est l'exposant de ramifica-

[22] Voir *Thèse*, Chap. I, p. 375.

[23] C'est le théorème d'existence de la base normale. Voir par exemple *Thèse*, Chap. IV, p. 397.

[24] Voir *Thèse*, Chap. V, p. 416.

[25] Par $Sp_{Z/K}X$ nous entendons la trace de $X$ prise de $Z$ à $K$.

tion de $\mathfrak{P}$ par rapport à $K$, les ordres pour $\mathfrak{P}$ des nombres de $K$ sont les multiples de $e$, d'où $[Z^*:K^*U_Z] = e$ et $[K^*:N_{Z/K}Z^*] = ef = [Z:K]$, ce qui démontre notre formule.

Si $\mathfrak{p}$ est un diviseur premier infini réel, et si $Z \neq K$, les normes relatives des nombres de $Z$ sont les nombres positifs de $K$, de sorte que $[K^*:N_{Z/K}Z^*] = 2 = [Z:K]$. Dans tous les autres cas, la formule est trivialement évidente.

REMARQUE. Revenant au cas où $\mathfrak{p}$ est un diviseur premier fini, la formule $[U_K:N_{Z/K}U_Z] = e$ nous donne le résultat suivant:

*$K$ étant un corps de nombres algébriques de degré fini, $\mathfrak{p}$ un diviseur premier de $K$, $Z/K$ une extension cyclique de $K$, une condition nécessaire et suffisante pour que $U_\mathfrak{p} \subset N_{Z/K}J_Z$ est que $\mathfrak{p}$ ne soit pas ramifié dans $Z$.*

## 7. Limitation inferieure du nombre des différentielles attachees a une extension cyclique

Soit $Z/K$ une extension cyclique de degré $n$ du corps de nombres algébriques $K$. Si $E$ désigne un ensemble fini de diviseurs premiers de $K$, nous désignerons par $E'$ l'ensemble des diviseurs premiers de $Z$ qui divisent les diviseurs premiers de $E$. On peut choisir $E$ de telle manière que $J_K = P_K J_K^E$ et $J_Z = P_Z J_Z^{E'}$.[26] Nous supposerons de plus que $E$ contient tous les diviseurs premiers infinis de $K$, ainsi que les diviseurs premiers ramifiés dans $Z$.

Le nombre des différentielles de $K$ associées à l'extension $Z/K$ est l'indice (fini) $[J_K:P_K N_{Z/K}J_Z]$. Cet indice s'écrit

$$[P_K J_K^E : P_K N_{Z/K} J_Z^{E'}] = [J_K^E : N_{Z/K} J_Z^{E'}] \cdot [J_K^E \cap P_K N_{Z/K} J_Z^{E'} : N_{Z/K} J_Z^{E'}]^{-1}$$

Le premier facteur au second membre est égal à $\prod_{\mathfrak{p} \in E} [K_\mathfrak{p}^* : K_\mathfrak{p}^* \cap N_{Z/K}J_Z]$. En effet, si $\mathfrak{q} \notin E$, la $\mathfrak{q}$-coordonnée d'un élément de $J_K^E$, qui est une unité $\mathfrak{q}$-adique, est norme relative d'une unité de $ZK_\mathfrak{q}$, en vertu de la remarque terminale du §6. Si $n_\mathfrak{p} = [ZK_\mathfrak{p}:K_\mathfrak{p}]$ est l'ordre du groupe de décomposition de $\mathfrak{p}$ dans l'extension $Z/K$, on a donc $[J_K^E:N_{Z/K}J_Z^{E'}] = \prod_{\mathfrak{p} \in E} n_\mathfrak{p}$.

On a

$$[J_K^E \cap P_K N_{Z/K} J_Z^{E'} : N_{Z/K} J_Z^{E'}] = [P_K^E N_{Z/K} J_Z^{E'} : N_{Z/K} J_Z^{E'}]$$
$$= [P_K^E : P_K^E \cap N_{Z/K} J_Z^{E'}] = [P_K^E : N_{Z/K} P_Z^{E'}] \cdot [P_K^E \cap N_{Z/K} J_Z^{E'} : N_{Z/K} P_Z^{E'}]^{-1}.$$

Nous ne calculerons que le premier des facteurs du dernier membre de cette égalité. Il nous faudra pour cela introduire un sous-groupe d'indice fini de $P_Z^{E'}$. Soient $\mathfrak{p}_0, \mathfrak{p}_1, \cdots, \mathfrak{p}_s$ les diviseurs premiers de $E$. Pour chaque $i$ $(0 \leqq i \leqq s)$ choisissons un diviseur premier $\mathfrak{P}_i$ de $\mathfrak{p}_i$ dans $Z$ ainsi qu'un nombre $\mathsf{H}_i^{**} \in P_Z^{E'}$ tel que $w_{\mathfrak{P}_i}(\mathsf{H}_i^{**}) > 1$, $w_\mathfrak{Q}(\mathsf{H}_i^{**}) < 1$ pour tous les $\mathfrak{Q}$ de $E'$ autres que $\mathfrak{P}_i$ ($w_\mathfrak{P}$ représentant une valuation $\mathfrak{P}$-adique de $Z$).[27] Soit $\mathsf{H}_i^*$ la norme relative de $\mathsf{H}_i^{**}$ par rapport au corps de décomposition $Z_i$ de $\mathfrak{p}_i$ dans l'extension $Z/K$: $\mathsf{H}_i^*$ satis-

[26] Voir Corollaire 1, §3, p. 401.

[27] Voir Théorème 3, §3, p. 402.

fait encore aux mêmes conditions que $\mathsf{H}_i^{**}$. Tout diviseur premier de $E'$ est de la forme $\mathfrak{P}_i^\tau$, $\tau$ appartenant au groupe de Galois de $Z/K$, et on a $w_{\mathfrak{P}_i^\tau}((\mathsf{H}_i^*)^\tau) > 1$, $w_{\mathfrak{Q}}((\mathsf{H}_i^*)^\tau) < 1$ pour les $\mathfrak{Q}$ de $E'$ autres que $\mathfrak{P}_i^\tau$. Il en résulte que les $\mathsf{H}_i^*$ et leurs conjugués par rapport à $K$ engendrent un sous-groupe d'indice fini de $P_Z^{E'}$. Posons encore $\eta_i = N_{Z_i/K}\,\mathsf{H}_i^*$ et $\mathsf{H}_i = (\mathsf{H}_i^*)^{n_{\mathfrak{p}_i}}\eta_i^{-1}$; le groupe engendré par $\eta_0$, $\eta_1, \cdots, \eta_s$, par les $\mathsf{H}_i$ et par les conjugués de ces derniers est encore d'indice fini dans $P_Z^{E'}$, car ce groupe contient les $(\mathsf{H}_i^*)^n$. D'autre part, les $s+1$ éléments $\eta_i$ sont dans $P_K^E$; comme il n'y a pas plus de $s$ éléments indépendants dans $P_K^E$,[27] il existe au moins une relation entre les $\eta_i$; supposant que, dans cette relation, $\eta_0$ figure avec un exposant $\neq 0$, on voit que le groupe $\Phi$ engendré par $\eta_1, \eta_2, \cdots, \eta_s$, les $\mathsf{H}_i$ et leurs conjugués est d'indice fini dans $P_Z^{E'}$. En tenant compte des relations $\mathsf{H}_i \in Z_i$, $N_{Z_i/K}\,\mathsf{H}_i = 1$, on voit que $\Phi$ est encore engendré par les $s + \sum_{i=0}^{s}(m_i - 1) = \sum_{i=0}^{s} m_i - 1$ éléments $\eta_1, \eta_2, \cdots, \eta_s$, $\mathsf{H}_i^{\sigma^k}$ $(0 \leqq k < m_i - 1)$, où $m_i$ représente le nombre $n/n_{\mathfrak{p}_i}$ et où $\sigma$ est un générateur du groupe de Galois de $Z/K$. Mais $\sum_{i=0}^{s} m_i$ est le nombre des diviseurs premiers de $E'$; il en résulte qu'il n'y a aucune relation entre les générateurs que nous venons d'indiquer. Tout élément de $\Phi$ se met sous la forme $\prod_{i=1}^{s}\eta_i^{u_i}\prod_{i=0}^{s}\mathsf{H}_i^{f_i(\sigma)}$ les $u_i$ étant des entiers et les $f_i(\sigma)$ des polynomes à coefficients entiers déterminés modulo $1 + \sigma + \cdots + \sigma^{m_i-1}$.

Ceci posé, les applications $\mathsf{E} \to N_{Z/K}\,\mathsf{E}$, $\mathsf{E} \to \mathsf{E}^{1-\sigma}$ sont des homomorphismes $L_1$, $L_2$ de $P_Z^{E'}$ dans lui-même, et on a $L_1L_2\,\mathsf{E} = L_2\,L_1\,\mathsf{E} = 1$. Le groupe $\Phi$ est appliqué dans lui-même par ces homomorphismes. Désignons par $Q$ le groupe des nombres de $P_Z^{E'}$ de norme relative 1 par rapport à $K$ et par $R$ celui des nombres $\mathsf{A}$ de $P_Z^{E'}$ tels que $\mathsf{A}^{1-\sigma} = 1$. Le lemme de Herbrand[28] nous donne la formule

$$\frac{[R : N_{Z/K}P_Z^{E'}]}{[Q : (P_Z^{E'})^{1-\sigma}]} = \frac{[R \cap \Phi : N_{Z/K}\Phi]}{[Q \cap \Phi : \Phi^{1-\sigma}]}$$

Or $R$ est évidemment égal à $P_K^E$. Je dis que le dénominateur du premier membre est égal à 1. En effet, si $\mathsf{A} \in Q$, il existe un nombre $\mathsf{B} \in Z$ tel que $\mathsf{A} = \mathsf{B}^{1-\sigma}$; $\mathfrak{Q}$ étant un diviseur premier ne figurant pas dans $E'$, $\mathfrak{Q}$ n'est pas ramifié par rapport à $K$, de sorte que tout nombre de $Z_{\mathfrak{Q}}^*$ est le produit d'un nombre de $K_{\mathfrak{q}}^*$ par une unité $\mathfrak{Q}$-adique ($\mathfrak{q}$ désignant le diviseur premier de $K$ divisible par $\mathfrak{Q}$). On a donc $J_Z = J_K\,J_Z^{E'} = P_K\,J_K^E\,J_Z^{E'} = P_K\,J_Z^{E'}$, et aussi $P_Z = P_K\,P_Z^{E'}$; le nombre $\mathsf{B}$ se met donc sous la forme $\beta\Gamma$, où $\beta \in P_K$, $\Gamma \in P_Z^{E'}$; d'où $\mathsf{A} = \Gamma^{1-\sigma}$, ce qui démontre notre assertion.

Nous allons maintenant calculer les indices au second membre de notre formule. Soit $\mathsf{A} = \prod_{i=1}^{s}\eta_i^{u_i}\prod_{i=0}^{s}\mathsf{H}_i^{f_i(\sigma)}$ un élément de $\Phi$. Si $\mathsf{A}^{1-\sigma} = 1$, on doit avoir $f_i(\sigma)(1-\sigma) \equiv 0 \pmod{1 + \sigma + \cdots + \sigma^{m_i-1}}$, ce qui entraîne $f_i(\sigma)$

[28] Voir *Thèse*, Chap. I, p. 375.

$\equiv 0 \pmod{1 + \sigma + \cdots + \sigma^{m_i - 1}}$. Donc $R \cap \Phi$ est le groupe engendré par les $\eta_i$ $(1 \leqq i \leqq s)$; $N_{Z/K}\Phi$ est le groupe engendré par les $\eta_i^n$ $(1 \leqq i \leqq s)$, d'où $[R \cap \Phi : N_{Z/K}\Phi] = n^s$.

Pour que $\mathsf{A} \in Q$, il faut et suffit que l'on ait $u_i = 0$ $(1 \leqq i \leqq s)$; pour que $\mathsf{A} \in \Phi^{1-\sigma}$, il faut et suffit que les $u_i$ soient nuls, et que de plus il existe des polynômes $g_i(\sigma)$ $(0 \leqq i \leqq s)$ à coefficients entiers tels que $f_i(\sigma) \equiv g_i(\sigma)(1 - \sigma) \pmod{1 + \sigma + \cdots + \sigma^{m_i - 1}}$; ces dernières conditions entraînent $f_i(1) \equiv 0 \pmod{m_i}$ $(0 \leqq i \leqq s)$; inversement, supposons $f_i(\sigma) = k_i m_i$, $k_i$ étant un entier; les polynômes $f_i(\sigma) - f_i(1)$ et $k_i(1 + \sigma \cdots + \sigma^{m_i - 1}) - k_i m_i$ étant divisibles par $1 - \sigma$, il existera un $g_i(\sigma)$ tel que $f_i(\sigma) \equiv g_i(\sigma)(1 - \sigma) \pmod{1 + \sigma + \cdots + \sigma^{m_i - 1}}$. Il résulte de là que l'indice $(Q \cap \Phi : \Phi^{1-\sigma})$ vaut $\prod_{i=0}^{s} m_i = n^{s+1} \left(\prod_{\mathfrak{p} \in E} n_{\mathfrak{p}}\right)^{-1}$.

Finalement, nous obtenons la formule

$$[J_K : P_K N_{Z/K} J_Z] = [Z:K] \cdot [P_K^E \cap N_{Z/K} J_Z^{E'} : N_{Z/K} P_Z^{E'}]$$

qui démontre que le nombre des différentielles associées à l'extension cyclique $Z/K$ est divisible par $[Z:K]$.

### 8. Ou sont effectues certains calculs preliminaires

1. $n$ étant un entier $> 0$, $K$ un corps contenant les racines $n$-ièmes de l'unité, soit $\mathfrak{p}$ un diviseur premier fini de $K$. Nous allons calculer l'indice $[K_{\mathfrak{p}}^* : (K_{\mathfrak{p}}^*)^n]$. Sois $\pi$ un élément d'ordre 1 pour $\mathfrak{p}$; un élément de $K_{\mathfrak{p}}^*$ se met sous la forme $\pi^i \mathfrak{u}$, avec $\mathfrak{u} \in U_{\mathfrak{p}}$ ; cet élément n'est puissance $n$-ième exacte dans $K_{\mathfrak{p}}^*$ que si $i \equiv 0 \pmod{n}$ et si $\mathfrak{u} \in (U_{\mathfrak{p}})^n$. On a donc $[K_{\mathfrak{p}}^* : (K_{\mathfrak{p}}^*)^n] = n[U_{\mathfrak{p}} : (U_{\mathfrak{p}})^n]$. $M$ étant un entier $> 0$, soit $V$ le groupe des racines $M$-ièmes de l'unité dans $K_{\mathfrak{p}}$ ; on a $[U_{\mathfrak{p}} : (U_{\mathfrak{p}})^n] = [U_{\mathfrak{p}} : V(U_{\mathfrak{p}})^n][V(U_{\mathfrak{p}})^n : (U_{\mathfrak{p}})^n] = [(U_{\mathfrak{p}})^M : (U_{\mathfrak{p}})^{nM}] \cdot [V : V \cap (U_{\mathfrak{p}})^n]$. Si $M$ est multiple de $n$, le second facteur est égal à $n$, car $V$ est alors un groupe cyclique d'ordre multiple de $n$. Nous choisirons de plus $M$ tel que la fonction logarithmique[29] soit définie et univalente sur $(U_{\mathfrak{p}})^M$ et que le groupe additif $\log (U_{\mathfrak{p}})^M$ contienne tous les entiers divisibles par une certaine puissance de $\pi$. Soit alors $\mathfrak{M}$ ce groupe. Désignons par $p$ le nombre premier rationnel divisible par $\mathfrak{p}$, par $\mathfrak{o}_{\mathfrak{p}}$ l'anneau des entiers du corps $R_p$ des nombres $p$-adiques rationnels, par $\mathfrak{O}_{\mathfrak{p}}$ celui des entiers de $K_{\mathfrak{p}}$. $\mathfrak{M}$ et $\mathfrak{O}_{\mathfrak{p}}$ sont des modules par rapport à $\mathfrak{o}_{\mathfrak{p}}$ de même dimension, donc opérateurs-isomorphes. D'autre part la fonction logarithmique applique $(U_{\mathfrak{p}})^{nM}$ sur $n\mathfrak{M}$, de sorte que l'indice $[(U_{\mathfrak{p}})^M : (U_{\mathfrak{p}})^{nM}]$ est égal à $[\mathfrak{M} : n\mathfrak{M}]$, ou encore à $[\mathfrak{O}_{\mathfrak{p}} : n\mathfrak{O}_{\mathfrak{p}}]$. Mais ce dernier indice est la norme de l'idéal $n\mathfrak{O}_{\mathfrak{p}}$ ; il est donc égal à $(N\mathfrak{p})^{e_{\mathfrak{p}}}$ où $N\mathfrak{p}$ représente la norme absolue de $\mathfrak{p}$, et $e_{\mathfrak{p}}$ l'ordre de $n$ pour $\mathfrak{p}$. D'où la formule $[K_{\mathfrak{p}}^* : (K_{\mathfrak{p}}^*)^n] = n^2 N(\mathfrak{p}^{e_{\mathfrak{p}}})$.

Dans le cas où $\mathfrak{p}$ est un divisieur premier infini réel, l'indice $[K_{\mathfrak{p}}^* : (K_{\mathfrak{p}}^*)^n]$ est égal à 2 ou à 1 suivant que $n$ est pair ou impair. Cependant, il convient de remarquer que le corps $K$, contenant les racines $n$-ièmes de l'unité, ne peut avoir de diviseur premier infini réel que si $n = 2$. Nous pouvons donc dire que, pour tout diviseur premier infini réel $\mathfrak{p}$, on a $[K_{\mathfrak{p}}^* : (K_{\mathfrak{p}}^*)^n] = n$.

[29] Voir *Thèse*, Chap. V, p. 416.

Enfin, si $\mathfrak{p}$ est infini imaginaire, $[K_\mathfrak{p}^* : (K_\mathfrak{p}^*)^n] = 1$.

REMARQUE 1. Si $\mathfrak{p}$ est un diviseur premier fini ne divisant pas $n$, les calculs faits plus haut montrent que $[U_\mathfrak{p} : (U_\mathfrak{p})^n] = n$.

2. $K$ étant toujours supposé contenir les racines $n$-ièmes de l'unité, soit $E$ un ensemble fini de diviseurs premiers de $K$ contenant tous les diviseurs premiers infinis et tous ceux qui divisent $n$. L'indice $[J_K^E : J_K^{E,n}]$ vaut $\sum_{\mathfrak{p} \in E} [K_\mathfrak{p}^* : (K_\mathfrak{p}^*)^n] = n^{2s'+r} \prod_{\mathfrak{p} \in E} N(\mathfrak{p}^{e_\mathfrak{p}})$, où $s'$ désigne le nombre des diviseurs premiers finis de $E$ et $r$ le nombre des diviseurs premiers infinis réels de $K$. Or on a $\prod_{\mathfrak{p} \in E} N(\mathfrak{p}^{e_\mathfrak{p}}) = N(n) = n^d$ si $d$ est le degré absolu de $K$. Mais $d + r$ est le double du nombre des diviseurs premiers infinis de $K$. Comme ces derniers sont tous dans $E$, on a

$$[J_K^E : J_K^{E,n}] = n^{2(s+1)}$$

si $s + 1$ est le nombre des diviseurs premiers de $E$.

3. $E$ et $K$ ayant la même signification que dans 2), soit $T$ le corps obtenu par adjonction à $K$ des racines $n$-ièmes de tous les nombres de $P_K^E$. On a $[T : K] = [P_K^E : (P_K^E)^n]$.[30] Or le groupe $P_K^E$ est le produit direct du groupe des racines de l'unité de $K$, qui est d'ordre multiple de $n$, et d'un groupe engendré par $s$ éléments indépendants. L'indice $[P_K^E : (P_K^E)^n]$ vaut donc $n^{s+1}$. D'où la formule $[T : K] = n^{s+1}$.

REMARQUE 2. $\mathfrak{p}$ étant un diviseur premier fini de $K$ ne divisant pas $n$, si $\alpha$ est un nombre de $K$ qui est dans $U_\mathfrak{p}$, $\mathfrak{p}$ n'est pas ramifié dans l'extension $K(\alpha^{1/n})/K$, car le discriminant du polynome $x^n - \alpha$ est $\alpha^{n-1}(-n)^n$ qui est une unité $\mathfrak{p}$-adique. Il en résulte que seuls les diviseurs premiers de $E$ peuvent être ramifiés dans $T$.

## 9. Limitation superieure du nombre des differentielles associées a une extension abélienne

$Z/K$ étant une extension abélienne finie du corps $K$, nous désignerons par $\Phi(K, Z)$ le groupe des differentielles de $K$ qui lui sont associées. L'objet de ce paragraphe est de démontrer que $[\Phi(K, Z) : 1]$ divise $[Z : K]$.

Soit $Z'$ un corps tel que $K \subset Z' \subset Z$. L'application $\varphi \to N_{K/Z'}\varphi$ est un homomorphisme de $\Phi(K, Z)$ dans $\Phi(Z', Z)$; les differentielles qui sont appliquées sur 1 par cet homomorphisme sont celles de $\Phi(K, Z')$. On en déduit que $[\Phi(K, Z) : 1]$ divise $[\Phi(Z', Z) : 1] \cdot [\Phi(K, Z') : 1]$. Par suite, si le résultat annoncé est vrai pour les extensions $Z'/K$ et $Z/Z'$, il est vrai pour $Z/K$. Il suffira donc de le démontrer pour une extension $Z/K$ de degré premier $n$. Soit alors $K'$ le corps obtenu par adjonction à $K$ des racines $n$-ièmes de l'unité. Nous savons que $[\Phi(K, ZK') : 1]$ divise $[\Phi(K', ZK') : 1] \cdot [\Phi(K, K') : 1]$. Or le degré $d$ de $K'/K$ divise $n - 1$; l'ordre d'un élément de $\Phi(K, K')$ étant évidemment un diviseur de $d$, $[\Phi(K, K') : 1]$ est premier à $n$. De plus $[\Phi(K, Z) : 1]$ est une

[30] Voir *Thèse*, Chap. IV, p. 394.

puissance de $n$ et divise $[\Phi(K, ZK'):1]$. Il en résulte immédiatement qu'il suffira de prouver notre assertion pour l'extension $ZK'/K'$.

Considérons donc le cas d'un corps $K$ contenant les racines $n$-ièmes de l'unité ($n$ étant un nombre premier) et d'une extension $Z/K$ de degré $n$. On a $Z = K(\beta_0^{1/n})$, $\beta_0$ étant un nombre de $K$. Soit $E$ un ensemble fini de diviseurs premiers de $K$, contenant tous les diviseurs premiers de $n$, tous les diviseurs premiers infinis, tous les diviseurs premiers ramifiés dans $Z$ et tel enfin que $J_K = P_K J_K^E$.[31] Nous pouvons supposer $\beta_0 \in P_K^E$. Donc $Z$ est contenu dans le corps $T$ obtenu par adjonction à $K$ des racines $n$-ièmes de tous les nombres de $P_K^E$. Si $s+1$ désigne le nombre des éléments de $E$, on peut choisir $s$ nombres $\beta_1, \beta_2, \cdots, \beta_s$ de $P_K^E$ tel que $T$ soit engendré par les $s+1$ nombres $\beta_i^{1/n} (0 \leqq i \leqq s)$. Soit $T_i$ le corps composé des $K(\beta_j^{1/n})$, pour $j \neq i$ $(0 \leqq i \leqq s)$. On a $[T:T_i] = n$. On peut donc trouver un diviseur premier $\mathfrak{P}_i$ de $T_i$ qui reste premier dans $T$ et qui divise un diviseur premier $\mathfrak{p}_i$ de $K$ non situé dans $E$.[32] Donc $\mathfrak{p}_i$ n'est pas ramifié dans $T/K$; son groupe de décomposition étant cyclique se confond avec le groupe de Galois de $T/T_i$. On a donc $K_{\mathfrak{p}_j}(\beta_i^{1/n}) = K_{\mathfrak{p}_j}$ si $j \neq i$, de sorte que $\beta_i$ est contenu dans $(K_{\mathfrak{p}_j}^*)^n$.

Soit $\mathfrak{a}_i$ un idèle de $K$, primaire et d'ordre 1 pour $\mathfrak{p}_i$. Nous allons montrer qu'un nombre $\theta \in K$ ne peut se mettre sous la forme $\prod_{i=0}^{s} \mathfrak{a}_i^{u_i} \cdot \mathfrak{b}$, avec $\mathfrak{b} \in J_K^{E,n}$ que si $\theta$ est une puissance $n$-ième dans $K$. En effet, posant $V = K(\theta^{1/n})$, soit $\varphi$ une différentielle associée à l'extension $V/K$. Si $\mathfrak{q}$ est un diviseur premier de $K$, et si $\mathfrak{q} \in E$, $\theta$ est une puissance $n$-ième dans $K_\mathfrak{q}^*$, de sorte que $\varphi(K_\mathfrak{q}^*) = \{1\}$. Si $\mathfrak{q}$ n'est pas dans $E$ et diffère de tous les $\mathfrak{p}_i$, $\mathfrak{q}$ n'est pas ramifié dans $V$ et on a $\varphi(U_\mathfrak{q}) = \{1\}$. D'autre part, on a $1 = \varphi(\beta_i) = \prod_{j=0}^{s} \varphi_{\mathfrak{p}_j}(\beta_i)$ où $\varphi_\mathfrak{q}$ désigne le $\mathfrak{q}$-composant de $\varphi$. Si $j \neq i$, la formule $\beta_i \in (K_{\mathfrak{p}_j}^*)^n$ entraîne $\varphi_{\mathfrak{p}_j}(\beta_i) = 1$; on a donc $\varphi_{\mathfrak{p}_i}(\beta_i) = 1$. Mais $\beta_i$ n'est pas dans $(U_{\mathfrak{p}_i})^n$ et nous savons que $[U_{\mathfrak{p}_1}:(U_{\mathfrak{p}_i})^n] = n$. Le groupe $U_{\mathfrak{p}_i}$ est donc engendré par $(U_{\mathfrak{p}_i})^n$ et par $\beta_i$. Il en résulte que $\varphi$ est identiquement égale à 1 sur ce groupe. Il résulte de tout ceci que $\varphi$ est identiquement égale à 1 sur $J_K^E$, donc aussi sur $J_K = P_K J_K^E$; donc que $\varphi = 1$. On a donc $V = K$, ce qui prouve notre assertion.

Nous en déduirons d'abord que l'on a $P_K^E \cap J_K^{E,n} = (P_K^E)^n$, et ensuite que, en posant $H = P_K J_K^{E,n}$, le groupe $H$ est d'indice $n^s$ dans le groupe engendré par $H$ et par $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_s$. Or, on a $[J_K:H] = [P_K J_K^E:P_K J_K^{E,n}] = [J_K^E:P_K^E J_K^{E,n}] = [J_K^E:J_K^{E,n}][P_K^E J_K^{E,n}:J_K^{E,n}] = n^{2(s+1)} n^{-(s+1)} = n^{s+1}$. D'ou $[J_K:H'] = n$. Or $H'$ est évidemment contenu dans le groupe $P_K N_{Z/K} J_Z$. Donc $[J_K:P_K N_{Z/K} J_Z]$ divise $n$, ce qui achève la démonstration du résultat énoncé au début de ce paragraphe.

[31] Voir Corollaire 1, §3, p. 401.

[32] Si tous les diviseurs premiers de $T_i$ sauf un nombre fini étaient complètement décomposés dans $T$, le groupe $N_{T/T_i} J_T$ contiendrait tous les groupes $(T_i)_\mathfrak{Q}^*$ sauf un nombre fini, en vertu de la formule démontrée au §6 ($\mathfrak{Q}$ parcourant tous les diviseurs premiers de $T_i$). D'après le lemme du §4, p. 403, il en résulterait $P_{T_i} N_{T/T_i} J_T = J_{T_i}$, ce qui est en contradiction avec le résultat du §6.

## 10. Les caractères locaux non ramifiés

Soit $\mathfrak{p}$ un diviseur premier fini du corps $K$, et soit $\chi$ un caractère de $\mathfrak{G}_K$, non ramifié en $\mathfrak{p}$. On sait qu'il existe dans le groupe de Galois de $Z_\chi/K$ un élément $\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)$, appelé substitution de Frobenius attachée à $\mathfrak{p}$, tel que l'on ait

$$\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)\theta \equiv \theta^{N\mathfrak{p}} \pmod{\mathfrak{p}}$$

pour tout entier $\theta$ de $Z_\chi$ ($N\mathfrak{p}$ désignant la norme absolue de $\mathfrak{p}$). $\mathfrak{a}$ étant un idèle de $K$, nous poserons

$$\left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right) = \chi\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)^\nu$$

où $\nu$ est l'ordre de $\mathfrak{a}$ pour $\mathfrak{p}$. La fonction de $\mathfrak{a}$ ainsi définie sera notée, si on ne désire pas préciser la valeur de l'argument, par $\left(\frac{*, \chi}{\mathfrak{p}}\right)$. $\left(\frac{*, \chi}{\mathfrak{p}}\right)$ est un caractère local non ramifié de $J_K$. L'ordre de $\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)$ étant égal à celui, $n_\mathfrak{p}$, du groupe de décomposition de $\mathfrak{p}$ dans l'extension $Z_\chi/K$, $\left(\frac{*, \chi}{\mathfrak{p}}\right)$ est un caractère d'ordre $n_\mathfrak{p}$. Il en résulte que tout caractère local non ramifié $\varphi_\mathfrak{p}$ de $J_K$ attaché à $\mathfrak{p}$ dont l'ordre divise $n_\mathfrak{p}$ est une puissance de $\left(\frac{*, \chi}{\mathfrak{p}}\right)$. En effet, soit $\pi$ un élément de $K$ d'ordre 1 pour $\mathfrak{p}$; $\varphi_\mathfrak{p}(\pi)$ est une racine $n_\mathfrak{p}$-ième de l'unité, et $\left(\frac{\pi, \chi}{\mathfrak{p}}\right)$ est une racine primitive $n_\mathfrak{p}$-ième de l'unité; il existe donc un exposant $a$ tel que $\varphi_\mathfrak{p}(\pi) = \left(\frac{\pi, \chi}{\mathfrak{p}}\right)^a$, et on a alors $\varphi_\mathfrak{p} = \left(\frac{*, \chi}{\mathfrak{p}}\right)^a$. $\chi'$ étant un autre caractère de $\mathfrak{G}_K$ non ramifié en $\mathfrak{p}$, $\mathfrak{p}$ n'est pas ramifié dans l'extension $Z_\chi Z_{\chi'}/K$, et $\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)$ est l'automorphisme de $Z_\chi/K$ induit par $\left(\frac{Z_\chi Z_{\chi'}/K}{\mathfrak{p}}\right)$. On a donc $\left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right) = \chi\left(\frac{Z_\chi Z_{\chi'}/K}{\mathfrak{p}}\right)^\nu$ et de même $\left(\frac{\mathfrak{a}, \chi'}{\mathfrak{p}}\right) = \chi'\left(\frac{Z_\chi Z_{\chi'}/K}{\mathfrak{p}}\right)^\nu$, $\left(\frac{\mathfrak{a}, \chi\chi'}{\mathfrak{p}}\right) = \chi\chi'\left(\frac{Z_\chi Z_{\chi'}/K}{\mathfrak{p}}\right)^\nu$. D'où la formule

$$\left(\frac{\mathfrak{a}, \chi\chi'}{\mathfrak{p}}\right) = \left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right)\cdot\left(\frac{\mathfrak{a}, \chi'}{\mathfrak{p}}\right).$$

Soit $\tau$ un isomorphisme de $K$ avec l'un de ses conjugués. Nous savons que $\tau$ définit un isomorphisme $s \to s^\tau$ de $\mathfrak{G}_K$ avec $\mathfrak{G}_{K^\tau}$, qu'on peut aussi considérer comme un isomorphisme entre les groupes de Galois de $Z_\chi/K$ et de $Z_{\chi^\tau}/K$. Il

est évident que les substitutions de Frobenius $\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)$ et $\left(\frac{Z_{\chi^r}/K}{\mathfrak{p}^r}\right)$ se correspondent dans cet isomorphisme. D'où la formule

$$\left(\frac{\mathfrak{a}^r, \chi^r}{\mathfrak{p}^r}\right) = \left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right).$$

Soit enfin $\Omega/K$ une extension finie quelconque de $K$. Nous voulons démontrer la formule

$$\left(\frac{N_{\Omega/K}\mathfrak{A}, \chi}{\mathfrak{p}}\right) = \prod_{\mathfrak{P}} \left(\frac{\mathfrak{A}, N_{K/\Omega}\chi}{\mathfrak{p}}\right)$$

le produit étant étendu aux diviseurs premiers $\mathfrak{P}$ de $\mathfrak{p}$ dans $\Omega$. On peut supposer sans restriction que $\mathfrak{A}$ est primaire pour l'un de ces diviseurs premiers $\mathfrak{P}$. On sait que $\mathfrak{P}$ n'est pas ramifié dans l'extension $Z_\chi\Omega/\Omega = Z_{N_{K/\Omega}\chi}/\Omega$. Soit $\nu_1$ l'ordre de $\mathfrak{A}$ pour $\mathfrak{P}$: le second membre de la formule vaut $(N_{K/\Omega}\chi)\left(\frac{\Omega Z_\chi/\Omega}{\mathfrak{P}}\right)^{\nu_1}$. Si $f$ est le degré relatif de $\mathfrak{P}$ par rapport à $K$, on a $N\mathfrak{P} = (N\mathfrak{p})^f$, et, pour tout entier $\theta$ de $Z_\chi$,

$$\left(\frac{\Omega Z_\chi/K}{\mathfrak{P}}\right)\theta \equiv \theta^{(N\mathfrak{p})^f} \equiv \left(\frac{Z_\chi/K}{\mathfrak{p}}\right)^f \theta \pmod{\mathfrak{P}}$$

de sorte que $\left(\frac{\Omega Z_\chi/K}{\mathfrak{P}}\right)$ induit dans $Z_\chi/K$ l'automorphisme $\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)^f$. Donc le second membre de la formule à démontrer vaut $\chi\left(\frac{Z_\chi/K}{\mathfrak{p}}\right)^{f\nu_1} = \left(\frac{N_{\Omega/K}\mathfrak{A}, \chi}{\mathfrak{p}}\right)$, car $N_{\Omega/K}\mathfrak{A}$ est d'ordre $f\nu_1$ pour $\mathfrak{p}$.

### 11. Sur les caractères Circulaires

Soit $\chi$ un caractère *circulaire* du corps $K$, c'est-à-dire tel que l'extension $Z_\chi/K$ soit contenue dans une extension de la forme $K(\zeta)/K$, $\zeta$ étant une racine de l'unité. Nous nous proposons d'individualiser l'une des différentielles associées à l'extension $Z_\chi/K$. Designant par $m$ l'ordre de $\zeta$, soit $E$ l'ensemble des diviseurs premiers de $K$ qui sont infinis ou qui divisent $m$. Nous allons montrer que, si $n$ est l'ordre de $\chi$, et si $\alpha$ est un nombre de $K$ appartenant à tous les $(K_\mathfrak{p}^*)^n$, pour $\mathfrak{p} \in E$, on a $\prod_{\mathfrak{p} \notin E} \left(\frac{\alpha, \chi}{\mathfrak{p}}\right) = 1$. (On remarquera que les diviseurs premiers non contenus dans $E$ ne sont pas ramifiés dans $Z_\chi$, ce qui montre que la formule écrite a un sens.) On peut trouver un nombre $\gamma$ tel que $\alpha\gamma^n$ soit un entier $\equiv 1 \pmod{m}$ et totalement positif.[33] On a alors $\prod_{\mathfrak{p} \notin E} \left(\frac{\alpha, \chi}{\mathfrak{p}}\right) = \prod_{\mathfrak{p} \notin E} \left(\frac{\alpha\gamma^n, \chi}{\mathfrak{p}}\right)$, de sorte qu'il suffit de démontrer l'hypothèse dans le cas où $\alpha$ est lui-même un entier $\equiv 1 \pmod{m}$ et totalement positif.

[33] Cela résulte du lemme du §4, p. 403.

Or, si $\nu_\mathfrak{p}$ est l'ordre pour $\mathfrak{p}$ de $\alpha$, on a $\left(\frac{\alpha, \chi}{\mathfrak{p}}\right) = \chi\left(\frac{K(\zeta)/K}{\mathfrak{p}}\right)^{\nu_\mathfrak{p}}$. Il suffira donc de montrer que l'élément $s = \prod_{\mathfrak{p} \notin E} \left(\frac{K(\zeta)/K}{\mathfrak{p}}\right)^{\nu_\mathfrak{p}}$ du groupe de Galois de $K(\zeta)/K$ est égal à 1, c'est-à-dire que $s\zeta = \zeta$. Pour chaque $\mathfrak{p} \notin E$, on a

$$\left(\frac{K(\zeta)/K}{\mathfrak{p}}\right)\zeta \equiv \zeta^{N\mathfrak{p}} \pmod{\mathfrak{p}}.$$

Or, les deux membres de cette congruence sont des puissances de $\zeta$; $\mathfrak{p}$ ne divisant pas $m$, cette congruence entraîne l'égalité des deux membres. D'où

$$s\zeta = \zeta^{\prod_{\mathfrak{p} \notin E} (N\mathfrak{p})^{\nu_\mathfrak{p}}} = \zeta^{|N\alpha|} = \zeta$$

car $N\alpha$ est un entier rationnel positif et $\equiv 1 \pmod{m}$.

$\mathfrak{a}$ désignant maintenant un idèle de $K$, il existe un nombre auxiliaire[34] $\beta$ tel que $(\mathfrak{a}\beta^{-1})_\mathfrak{p} \in (K_\mathfrak{p}^*)^n$ pour tous les $\mathfrak{p} \in E$. Nous poserons

$$\varphi_\chi(\mathfrak{a}) = \prod_{\mathfrak{p} \notin E} \left(\frac{\mathfrak{a}\beta^{-1}, \chi}{\mathfrak{p}}\right).$$

Le second membre ne dépend pas du choix du nombre auxiliaire $\beta$, car, si on le remplace par un autre, $\mathfrak{a}\beta^{-1}$ se trouve multiplié par un nombre qui appartient à tous les $(K_\mathfrak{p}^*)^n$ pour $\mathfrak{p} \in E$. La fonction $\varphi_\chi$ ainsi définie est évidemment une différentielle. Je dis qu'elle est associée à l'extension $Z_\chi/K$. En effet, $\mathfrak{A}$ étant un idèle de $Z_\chi$, on peut trouver un nombre $\mathsf{B} \in Z_\chi$ tel que $(\mathfrak{A}\mathsf{B}^{-1})_\mathfrak{P} \in ((Z_\chi)_\mathfrak{P}^*)^n$, pour tous les diviseurs premiers $\mathfrak{P}$ de $Z_\chi$ qui divisent des diviseurs premiers de $E$. Dans ces conditions $N_{Z_\chi/K}(\mathsf{B})$ est un nombre auxiliaire pour l'idèle $N_{Z_\chi/K}(\mathfrak{A})$, et on a

$$\varphi_\chi(N_{Z_\chi/K}\mathfrak{A}) = \prod_{\mathfrak{p} \notin E} \left(\frac{N_{Z_\chi/K}\mathfrak{A}\mathsf{B}^{-1}, \chi}{\mathfrak{p}}\right) = 1.$$

Il est clair que, si $\mathfrak{p} \notin E$, le $\mathfrak{p}$-composant de $\varphi_\chi$ est $\left(\frac{*, \chi}{\mathfrak{p}}\right)$. Je dis maintenant que l'ordre de $\varphi_\chi$ est égal à l'ordre $n$ de $\chi$. S'il n'en était pas ainsi, il existerait un diviseur $n' < n$ de $n$ tel que $\varphi_\chi^{n'} = 1$. On aurait donc $\left(\frac{\mathfrak{a}, \chi^{n'}}{\mathfrak{p}}\right) = 1$ pour tout $\mathfrak{a} \in J_K$ et tout $\mathfrak{p} \notin E$. Soit $Z'$ le sous-corps de $Z$ appartenant au caractère $\chi^{n'}$: tous les diviseurs premiers non contenus dans $E$ de $K$ seraient complètement décomposés dans $Z'$, ce qui est impossible.[35]

Nous dirons que la différentielle $\varphi_\chi$ que nous venons de construire est associée à $\chi$. Elle est entièrement caractérisée par la propriété suivante: pour presque

[34] On reconnait la méthode utilisée par H. Hasse pour définir le symbole de restes normiques.

[35] Voir note (32), §9, p. 411.

tous les diviseurs premiers $\mathfrak{p}$ de $K$, le $\mathfrak{p}$-composant de $\varphi_\chi$ est $\left(\frac{*, \chi}{\mathfrak{p}}\right)$. Il résulte de là, et des propriétés des caractères locaux non ramifiés, que:

I. L'application $\chi \to \varphi_\chi$ est un homomorphisme du groupe des caractères circulaires $\chi$ dans le groupe des différentielles de $K$. Cette application est même un isomorphisme, puisque l'ordre de $\chi$ est égal à celui de $\varphi_\chi$.

II. $\Omega/K$ étant une extension finie de $K$, $\chi$ un caractère circulaire de $\mathfrak{G}_K$, on a $\varphi_{N_{K/\Omega}\chi} = N_{K/\Omega}\varphi_\chi$ (On notera que $N_{K/\Omega}\chi$ est un caractère circulaire de $\mathfrak{G}_\Omega$)

Enfin, comme nous savons que le groupe des différentielles associées à $Z_\chi/K$ est d'ordre $n$ égal à celui de $\chi$, il en résulte que ce groupe est cyclique, et qu'il est engendré par $\varphi_\chi$.

Nous aurons encore besoin du théorème d'existence suivant:

THÉORÈME 4. *Soient $\mathfrak{p}$ un diviseur premier fini de $K$, $n$ un entier $> 0$, $\Omega/K$ une extension finie de $K$. Il existe un caractère circulaire $\chi$ de $\mathfrak{G}_K$ non ramifié en $\mathfrak{p}$ tel que $\left(\frac{*, \chi}{\mathfrak{p}}\right)$ soit d'ordre $n$ et que $Z_\chi \cap \Omega = K$.*

Désignons par $f$ le degré absolu de $\mathfrak{p}$, par $p$ le nombre premier rationnel divisible par $\mathfrak{p}$, par $R$ le corps des rationnels. Il nous suffira de prouver l'existence d'un caractère circulaire $\tilde{\chi}$ de $\mathfrak{G}_R$ non ramifié en $p$ tel que $\left(\frac{*, \tilde{\chi}}{p}\right)$ soit d'ordre $fn$ et que $Z_{\tilde{\chi}} \cap \Omega = R$, car alors $\chi = N_{R/K}\tilde{\chi}$ satisfera aux conditions imposées.

Désignons par $m$ un entier divisible par les ordres de toutes les racines de l'unité contenues dans $\Omega$.

$v$ étant un nombre premier, soit $v^a$ la contribution de $v$ à $fn$. $q$ étant un nombre premier $\neq p$, l'ordre de la substitution de Frobenius $\left(\frac{R(e^{2i\pi/q})/R}{p}\right)$ est le plus petit exposant $x$ tel que $p^x \equiv 1 \pmod q$. Posons $S_h = p^{v^h} - 1$, $Q_h = S_{h+1}/S_h$. On a

$$Q_h = 1 + p^{v^h} + \cdots + p^{(v-1)v^h} \equiv v \pmod{S_h} \qquad Q_h > v.$$

Donc, le p.g.c.d. de $Q_h$ et de $S_h$ divise $v$. Si $p \not\equiv 1 \pmod v$, on a aussi $S_h \not\equiv 0 \pmod v$ pour tout $h$. Si $p \equiv 1 \pmod v$, on a $Q_0 \equiv 0 \pmod v$ et $S_h \not\equiv 0 \pmod{v^2}$ pour $h \geq 1$. Par suite, si $h \geq 1$, $Q_h$ n'est pas divisible par $v^2$ et possède un facteur premier $q_{v,h}$ qui ne divise pas $S_h$, mais qui divise les $S_{h'}$ pour $h' > h$. Il en résulte qu'on peut trouver un $h > a$ tel que $q_v = q_{v,h}$ soit premier à $m$. L'ordre de $\left(\frac{R(e^{2i\pi/q_v})/R}{p}\right)$ sera $v^h$; on pourra donc trouver un caractère $\tilde{\chi}_v$ du groupe de Galois de $R(e^{2i\pi/q_v})/R$ tel que $\left(\frac{*, \tilde{\chi}_v}{p}\right)$ soit un caractère d'ordre $v^a$. Le caractère $\tilde{\chi} = \prod_{v/fn} \tilde{\chi}_v$ satisfera a toutes les conditions imposées.

## 12. Comparaison avec des extensions circulaires

Soient $\chi$ un caractère de $\mathfrak{G}_K$, d'ordre $n$, et $\varphi$ une différentielle associée à $Z_\chi/K$. Soit $\chi'$ un caractère circulaire de $\mathfrak{G}_K$, d'ordre $\equiv 0 \pmod n$, indepen-

dant[36] de $\chi$; désignons par $\varphi'$ la différentielle associée à $\chi'$. $a$ étant un entier, posons $T' = Z_\chi \chi'^a$. Donc $N_{K/T'}\chi = N_{K/T'}\chi'^a$, de sorte que l'extension $Z_\chi T'/T'$ est circulaire. La différentielle associée à $N_{K/T'}\chi'^{-a}$ est $N_{K/T'}\varphi'^{-a}$; comme $N_{K/T'}\varphi$ est une différentielle associée à l'extension $Z_\chi T'/T'$, on en déduit qu'il existe un entier $u$ tel que $N_{K/T'}\varphi = N_{K/T'}\varphi'^u$. Nous voulons maintenant montrer que $u$ est proportionnel à $a$ (mod $n$), c'est-à-dire qu'il existe un entier $v$, indépendant de $a$, tel que $u \equiv av \pmod n$.

Pour cela, introduisons un caractère circulaire $\chi''$ de $\mathfrak{G}_K$, d'ordre $\equiv 0 \pmod n$, tel que $\chi$, $\chi'$, $\chi''$ soient indépendants. Désignons par $T''$ le corps $Z_{\chi\chi''}$, et par $\varphi''$ la différentielle associée à $\chi''$. En opérant comme il y a un instant, on prouve que $N_{K/T''}\varphi$ se met sous le forme $N_{K/T''}\varphi''^v$.

Considerons maintenant l'extension $T'T''/T''$. La différentielle $\varphi\varphi'^{-u}$ étant associée à $T'/K$, il en resulte que $N_{K/T''}\varphi\varphi'^{-u} = N_{K/T''}\varphi''^v\varphi'^{-u}$ est associée à $T'T''/T''$. Mais $N_{K/T''}\varphi''^v\varphi'^{-u}$ est la différentielle associée au caractère circulaire $N_{K/T''}\chi''^v\chi'^{-u}$. D'autre part, le caractère circulaire $N_{K/T''}\chi\chi'^a = N_{K/T''}\chi''^{-1}\chi'^a$ est un caractère primitif du groupe de Galois de $T'T''/T''$. On en conclut que $N_{K/T''}\chi''^v\chi'^{-u}$ se met sous la forme $(N_{K/T''}\chi''^{-1}\chi'^a)^x$ de sorte que

$$\chi''^v\chi'^{-u} = (\chi''^{-1}\chi'^a)^x(\chi\chi'')^y$$

d'où $-u \equiv ax$, $v \equiv -x + y$, $0 \equiv y \pmod n$, et $u \equiv av \pmod n$ ce qui démontre notre assertion.

Soit maintenant $\mathfrak{p}$ un diviseur premier fini de $K$ non ramifié dans $Z$. On peut choisir le caractère circulaire $\chi'$ de $\mathfrak{G}_K$ tel que $\varphi'_\mathfrak{p}$ soit un caractère non ramifié d'ordre $n$.[37] $\left(\frac{*,\ \chi}{\mathfrak{p}}\right)$ étant aussi un caractère non ramifié de $K_p^*$, dont l'ordre divise $n$, pourra se mettre sous la forme $(\varphi'_\mathfrak{p})^{-a}$, $a$ étant un entier. On aura donc $\left(\frac{*,\ \chi\chi'^a}{\mathfrak{p}}\right) = 1$, de sorte que, en conservant les notations que nous avions tout-à-l'heure, $T'K_\mathfrak{p} = K_\mathfrak{p}$. Par suite $(\varphi\varphi'^{-u}) = 1$, d'où $\varphi_\mathfrak{p} = \varphi'^{-u}_\mathfrak{p}$. Or on a $u \equiv av \pmod n$ et par suite $\varphi_\mathfrak{p} = \left(\frac{*,\ \chi'}{\mathfrak{p}}\right)^{-u} = \left(\frac{*,\ \chi'}{\mathfrak{p}}\right)^{-av} = \left(\frac{*,\ \chi}{\mathfrak{p}}\right)^v$.

Convenons pour un moment de dire que $v$ est un entier coordonné à la différentielle $\varphi$ associée à l'extension $Z_\chi/K$ si, pour presque tous les diviseurs premiers finis non ramifiés dans $Z_\chi$, on a $\varphi_\mathfrak{p} = \left(\frac{*,\ \chi}{\mathfrak{p}}\right)^v$. Si $v$, $v_1$ sont coordonnés aux différentielles $\varphi$, $\varphi_1$ et si $v \equiv v_1 \pmod n$, on a $\varphi = \varphi_1$.[38] Comme il y a $n$ différentielles distinctes associées à $Z_\chi/K$, il y en a une et une seule à laquelle soit

[36] Nous disons que des caractères $\chi$, $\chi'$, $\cdots$ sont indépendants si toute relation de la forme $\chi^a\chi'^{a'}\cdots = 1$, les $a$, $a'$, $\cdots$ étant des entiers, entraîne $\chi^a = 1$, $\chi'^{a'} = 1$, $\cdots$. Il revient au même de dire que l'intersection de chacun des corps $Z_\chi Z_{\chi'}$, $\cdots$ avec le corps composé de tous les autres se réduit au corps de base $K$.

[37] Voir Théorème 4, §11, p. 415.

[38] Car une différentielle est déjà bien determinée par la donnée de presque toutes ses coordonnées; voir §4, p. 403.

coordonné le nombre 1: c'est cette différentielle que nous appellerons la *différentielle associée* à $\chi$, et que nous désignerons par $\varphi_\chi$. De la formule $\left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right)\left(\frac{\mathfrak{a}, \chi_1}{\mathfrak{p}}\right) = \left(\frac{\mathfrak{a}, \chi\chi_1}{\mathfrak{p}}\right)$, valable pour un diviseur premier non ramifié dans $Z_\chi Z_{\chi_1}$, on déduit que $\varphi_{\chi\chi_1} = \varphi_\chi\varphi_{\chi_1}$ ; d'autre part, si $\varphi_\chi = 1$, on a, pour presque tous les $\mathfrak{p}$, $K_\mathfrak{p}Z_\chi = K_\mathfrak{p}$, d'où $Z_\chi = K$, $\chi = 1$. La correspondance $\chi \to \varphi_\chi$ est donc un isomorphisme du groupe des caractères de $\mathfrak{G}_K$ avec un groupe de différentielles de $K$.

L'isomorphisme $\chi \to \varphi_\chi$ jouit des propriétés suivantes:

1) Si $\tau$ est un isomorphisme de $K$ avec l'un de ses conjugués, on a $(\varphi_\chi)^\tau = \varphi_{\chi^\tau}$, comme il résulte de la formule $\left(\frac{\mathfrak{a}^\tau, \chi^\tau}{\mathfrak{p}^\tau}\right) = \left(\frac{\mathfrak{a}, \chi}{\mathfrak{p}}\right)$.

2) Si $\Omega/K$ est une extension finie, on a $\varphi_{N_{K/\Omega}\chi} = N_{K/\Omega}\varphi_\chi$, comme il résulte de la formule $\left(\frac{N_{K/\Omega}\mathfrak{A}, \chi}{\mathfrak{p}}\right) = \prod_\mathfrak{P}\left(\frac{\mathfrak{A}, \chi}{\mathfrak{P}}\right)$, le produit étant étendu aux diviseurs premiers $\mathfrak{P}$ de $\mathfrak{p}$ dans $\Omega$.

Enfin, on remarquera que le groupe des différentielles associées à une extension abélienne finie $T/K$, étant d'ordre $\leq [T:K]$ et contenant les différentielles associées aux caractères du groupe de Galois de $T/K$, qui sont en nombre $[T:K]$, est d'ordre $[T:K]$ et se compose des différentielles associées aux caractères du groupe de Galois de $T/K$.

### 13. Qui acheve la demonstration

Il ne nous reste plus maintenant qu'à prouver que, pour toute différentielle $\varphi$ d'un corps $K$, il existe un caractère $\chi$ de $\mathfrak{G}_K$ tel que $\varphi = \varphi_\chi$.

LEMME 1. *$K'$ étant une extension cyclique du corps $K$, soit $\varphi$ une différentielle de $K$ telle que $N_{K/K'}\varphi$ soit associée à un caractère $\chi'$ de $\mathfrak{G}_{K'}$. Alors $\varphi$ est associée à un caractère de $\mathfrak{G}_K$.*

En effet, $\tau$ étant un automorphisme de $K'$ par rapport à $K$, la différentielle associée à $\chi'^\tau$ est $(N_{K/K'}\varphi)^\tau = N_{K/K'}\varphi$; on a donc $\chi'^\tau = \chi'$. L'extension $K'/K$ étant cyclique, $\chi'$ peut se mettre sous la forme $N_{K/K'}\chi$, $\chi$ étant un caractère de $\mathfrak{G}_K$.[39] On a donc $N_{K/K'}\varphi\varphi_\chi^{-1} = 1$, ce qui prouve que $\varphi\varphi_\chi^{-1}$ est associée à un caractère $\chi_1$ du groupe de Galois de $K'/K$; d'où $\varphi = \varphi_\chi\varphi_{\chi_1} = \varphi_{\chi\chi_1}$.

LEMME 2. *Toute différentielle $\varphi$ d'un corps $K$ dont l'ordre $n$ est un nombre premier est associée à un caractère de $\mathfrak{G}_K$.*

1) Supposons que $K$ contienne les racines $n$-ièmes de l'unité. Soit $E$ un ensemble fini de diviseurs premiers de $K$ contenant tous les diviseurs premiers infinis, tous ceux qui divisent $n$ et ceux pour lesquels $\varphi$ est ramifiée, et de plus tel que $J_K = P_K J_K^E$. On a vu qu'en posant $H = P_K J_K^{E,n}$ on a: $[J_K:H] = n^{s+1} = [T:K]$[40] où $s+1$ est le nombre des diviseurs premiers de $E$, et $T$ le corps composé de toutes les extensions cycliques de degré $n$ de $K$ dans lesquelles seuls

[39] Voir Théorème 1, §1, p. 397.

[40] Voir §9, p. 411.

les diviseurs premiers de $E$ soient ramifiés. Le groupe des différentielles associées à l'extension $T/K$ est d'ordre $n^{s+1}$, et est contenu dans le groupe des différentielles égales à 1 sur $H$; or ce dernier est aussi d'ordre $n^{s+1}$. Donc la différentielle $\varphi$, qui est égale à 1 sur $H$, est associée à l'extension $T/K$, donc aussi à un caractère du groupe de Galois de cette extension.

2) Dans le cas général, soit $K'$ le corps obtenu par adjonction à $K$ des racines $n$-ièmes de l'unité: $K'/K$ est une extension cyclique, et $N_{K/K'}\varphi$ est une différentielle de $K'$ d'ordre égal à 1 ou à $n$: il suffit donc d'appliquer le lemme 1.

Nous pouvons maintenant achever la demonstration du théorème fondamental. $\varphi$ étant une différentielle d'un corps $K$, nous démontrerons par récurrence sur l'ordre $n$ de $\varphi$ que $\varphi$ est associée à un caractère de $\mathfrak{G}_K$. La proposition est vraie si $n = 1$; supposons la démontrée pour toutes les différentielles dont les ordres sont $< n$; si $p$ est un facteur premier de $n$, $\varphi^{n/p}$, qui est d'ordre premier $p$, est associée à un caractère $\chi_1$ de $\mathfrak{G}_K$ ; soit $K' = Z_{\chi_1}$ : la différentielle $N_{K/K'}\varphi$ est d'ordre $n/p < n$, et est par suite associée à un caractère de $\mathfrak{G}_{K'}$. Donc, en vertu du lemme 1, $\varphi$ est associée à un caractère de $\mathfrak{G}_K$.

The Institute For Advanced Study.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# RANK AND SPAN IN FUNCTIONAL TOPOLOGY

BY MARSTON MORSE

(Received August 9, 1939)

## 1. Introduction.

The analysis of functions $F$ on metric spaces $M$ of the type which appear in variational theories is made difficult by the fact that the critical limits, such as absolute minima, relative minima, minimax values etc., are in general infinite in number. These limits are associated with relative $k$-cycles of various dimensions and are classified as 0-limits, 1-limits etc. The number of $k$-limits suitably counted is called the $k^{\text{th}}$ type number $m_k$ of $F$. The theory seeks to establish relations between the numbers $m_k$ and the connectivities $p_k$ of $M$. The numbers $p_k$ are finite in the most important applications. It is otherwise with the numbers $m_k$.

The theory has been able to proceed provided one of the following hypotheses is satisfied. The critical limits cluster at most at $+\infty$; the critical points are isolated;[1] the problem is defined by analytic functions; the critical limits taken in their natural order are well-ordered. These conditions are not generally fulfilled. The generality of the theory rested upon the fact that the cases treated approximate in a certain sense the most general problems which it is desirable to treat. From this point of view it became necessary to study the way in which the critical points of the approximating function approximate[2] the critical points of the given function in number and in type. This has been followed with success in certain cases. It should be pursued much further but it becomes increasingly difficult.

As an example we refer to the study of minimal surfaces[3] of the topological type of the circular disc bounded by a fixed simple closed curve $g$. Even when $g$ is analytic and regular the number of minimal surfaces bounded by $g$ may be infinite as far as is known at the moment. No useful or simple conditions are known under which a minimal surface bounded by $g$ is isolated from other such minimal surfaces. Moreover a study of geodesics shows that the difficulties described here are inherent in all but the analytic case.

The present paper introduces a completely new way of meeting these difficulties. It is not by way of restrictive hypotheses. The hypotheses are those

[1] See for example the excellent book by Seifert and Threlfall, *Variationsrechnung im Grossen*, Berlin, Teubner (1939), or earlier papers of the author.

[2] See Morse, *The calculus of variations in the large*, American Mathematical Society Colloquium Publications 18, New York (1934). In particular Chapter VI.

[3] See Morse and Tompkins, *Minimal surfaces of general critical type*, Annals of Mathematics 40 (1939), pp. 443–472.

which are fulfilled in all of the general problems so far treated. The notion of the *span* of a relative cycle $u$ is introduced. It is the difference of two values of $F$ intrinsically associated with $u$. (See §8.) It measures the minimum length of an interval of values of $F$ essential to a description of the bounding relations of $u$. Only certain relative $k$-cycles termed $k$-caps are counted and among these $k$-caps only those whose spans exceed a fixed positive constant $e$. The number of $k$-caps of span $> e$ suitably counted is now denoted by $m_k^e$. The numbers $m_k^e$ are finite if the connectivities are finite, and the whole theory proceeds on a finite basis, obtaining relations between the type numbers $m_k^e$ and the connectivities $p_k$ similar to those which occur in the earlier analytic and non-degenerate cases. Except for being positive the number $e$ is arbitrary. Upon taking different values of $e$ one gets relations between different sets of $k$-limits.

A very special example may be helpful. Let the space $M$ be a simple closed curve in the $xy$-plane. On $M$ let $F = y$. If $F$ has a finite number of relative extrema on $M$ the number $m_0$ of relative minima equals the number $m_1$ of relative maxima. If $F$ has an infinite number of relative extrema this relation is apparently lost. It appears again however in the new theory in a more general form. The number $m_1^e$ of 1-caps with spans $> e$ equals the number $m_0^e$ of 0-caps with spans $> e$. Even when the relative extrema of $F$ are finite in number the relations $m_1^e = m_0^e$ contain more than the ordinary relation $m_1 = m_0$.

In general each $k$-cap $u$ possesses a "cap height" $a(u)$ such that $u$ is on the subset $F \leqq a$ of $M$ and on no subset $F \leqq b$ for which $b < a$. The function $a(u)$ is called a *rank function*.[4] It is one of a number of rank functions defined for subsets of chains or cycles. These rank functions satisfy four rank conditions whose investigation is a matter of abstract group theory. The two fundamental questions involved are the following. I. Let $A$ be an arbitrary set of $k$-caps $u$ over which a rank function $\rho(u)$ is defined. Let $g(\rho)$ be a maximal group of $k$-caps of $A$ with a fixed rank $\rho$. When is it true that a maximal group $g$ of $k$-caps of $A$ is a direct sum of the form

$$g = \sum_{\rho} g(\rho),$$

where $\rho$ ranges over all ranks of $k$-caps of $A$? II. When is it true that the dimension of the maximal group $g$ is independent of the choice of $g$ as maximal in $A$? Once these questions are answered the finiteness of the numbers $m_k^e$ is proven, and the relations between the numbers $m_k^e$ and $p_k$ are established.

The present paper is a theory of critical limits, not of critical points. A critical limit is always assumed by the function $F$ at some critical point provided certain conditions such as the compactness of the sets $F \leqq c$ and the upper-reducibility of $F$ are satisfied. This part of the theory is developed[5] in M1

[4] Rank functions of this type were introduced for the first time in Morse, *Functional topology and abstract variational theory*, Annals of Mathematics 38 (1937), pp. 386–449. We shall refer to this paper as M2.

[5] See Morse, *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques XCII, Paris, Gauthier-Villars (1938). We shall refer to this paper as M1.

and will be further extended in a later paper. The difficulties are not those of group theory but of analysis.

## 2. The rank conditions R

We shall be concerned with an additive abelian operator group $G$ with coefficients $\delta$ in a field $\Delta$. Let $[\rho]$ be a simply ordered set of elements. To certain elements $u$ of $G$ there shall be assigned an element $\rho(u)$ of $[\rho]$. The elements of $G$ for which $\rho(u)$ is not defined shall include the null element of $G$. We term $\rho(u)$ the *rank* of $u$.

Let $G(\rho, \omega)$ denote the subgroup of $G$ generated by the elements of $G$ with ranks at most $\omega$ using arbitrary coefficients $\delta$ in $\Delta$. The group $G(\rho, \omega)$ will in general contain non-null elements without rank. We shall assume that the rank function $\rho(u)$ satisfies rank conditions $R$ as follows:

I. *If $u$ has a rank and $\delta \neq 0$, $\rho(u) = \rho(\delta u)$.*

II. *If $u$, $v$ and $u + v$ have ranks, $\rho(u + v) \leqq \max [\rho(u), \rho(v)]$.*

III. *If $u$ and $v$ have different ranks, $u + v$ has a rank.*

IV. *If $u$ and $v$ are elements in $G(\rho, \omega)$ without rank, $\rho(u + v)$ cannot exist and equal $\omega$.*

Conditions I, II, III are independent as is readily seen. That condition IV is independent of conditions I, II, III is shown by the following example.

EXAMPLE 2.1. Let $G$ be generated by elements $a$, $b$, $c$ with $\Delta$ the field of integers mod 2. Suppose that $a$, $b$, $a + b$, $c$ have the rank 1, but that no other elements of $G$ have a rank. Rank conditions I, II, III are satisfied. The group $G(\rho, 1)$ is identical with $G$. Condition IV is not satisfied. The elements $a + c$ and $b + c$ have no rank while their sum $a + b$ has the rank 1.

EXAMPLE 2.2. Let $M$ be a compact metric space and $F$ a function which is continuous on $M$. Let $F_s$ be the subset of $M$ on which $F \leqq s$. Let $G$ be the group of all singular $k$-cycles on $M$ taken mod 2. Each non-bounding cycle $u$ of $G$ will be assigned a rank $t(u)$. Let $H$ be the homology class of $u$, and $v$ an arbitrary member of $H$. Let $b(v)$ be the maximum of $F$ on $v$. The greatest lower bound of the numbers $b(v)$ as $v$ ranges over $H$ will be called the *inferior cycle limit* of $u$ or of $H$ and will be termed the rank $t(u)$ of $u$. These ranks $t(u)$ satisfy the rank conditions I to IV.

The groups which we shall study are *operator* subgroups, that is subgroups which contain $\delta u$ whenever they contain $u$. Our isomorphisms will be operator isomorphisms, that is isomorphisms in which $\delta u$ corresponds to $\delta v$ whenever $u$ corresponds to $v$.

We shall be concerned with various classes of elements of $G$, for example, the class of elements having rank. A class $A$ of elements of $G$ will be termed an *operator* class if whenever $u$ is in $A$ and $\delta \neq 0$, $\delta u$ is in $A$. It will be convenient to exclude the null element from operator classes. A group $g$ in $A$ shall mean a subgroup of $G$ every element of which except the null element is in $A$. Such a group $g$ will be termed *maximal* in $A$ if $g$ is a proper subgroup of no subgroup of $G$ in $A$.

A *generating* set for a subgroup $g$ of $G$ is a set of elements of $g$ linear com-

binations of which, using coefficients in $\Delta$, include all elements of $g$. A generating set of $g$ no subset of which is a generating set of $g$ will be called a *base* for $g$. We shall be concerned with operator subgroups whose bases contain at most aleph null elements. It is easy to show that the cardinal number of elements in a base of such groups $g$ is the same for all bases of $g$. This number is called the *dimension* of $g$.

Let $A$ be an operator class of $G$. Maximal groups of elements in $A$ may have different dimensions. As an example let $A$ consist of the elements

$$a, \quad b, \quad a + b, \quad c,$$

in Example 2.1. The group generated by $a$ and $b$ and the group generated by $c$ are maximal in $A$, but these groups have different dimensions.

Any operator subgroup of $G$ all of whose elements except the null element have ranks $\rho$ will be termed a *$\rho$-subgroup* of $G$. We shall be concerned with the conditions under which any two maximal $\rho$-subgroups of $G$ have the same dimension. In this connection the concepts of $\rho$-class and $\rho$-isomorphism are important.

Two elements $u$ and $v$ of $G$ with the same rank $\rho$ will be said to be in the same *$\rho$-class* or rank class if $u - v$ has no rank or has a rank less than $\rho$. An isomorphism between two $\rho$-subgroups of $G$ will be termed a *$\rho$-isomorphism* if corresponding non-null elements are in the same $\rho$-class.

LEMMA 2.1. *In a $\rho$-subgroup of $G$ with a finite dimension $r$ the number of different ranks is at most $r$.*

The proof of this lemma is given in M1, p. 17. It depends merely on rank conditions I, II, III. The following lemma is a consequence of Lemma 4.1 of M1.

LEMMA 2.2. *Let $g$ be a subgroup of $G$ in an operator class $A$. A necessary and sufficient condition that $g$ be maximal in $A$ is that each element in $A$ have the form*

$$u = z + w,$$

*where $z$ is in $g$ and $w$ not in $A$.*

Two groups of elements with rank are not in general $\rho$-isomorphic if their dimensions are equal. The condition of $\rho$-isomorphism is thus much stronger than that of isomorphism. The significance of the following theorem will now be apparent.

THEOREM 2.1. *When rank conditions $R$ are satisfied, any two maximal groups of elements of $G$ with the same fixed rank $\rho$ are $\rho$-isomorphic.*

This theorem is proved on p. 15 of M1, and depends essentially upon all four rank conditions $R$. In this connection the following two theorems were also proved.

THEOREM 2.2. *Under rank conditions $R$ the property of elements with rank being in the same $\rho$-class is transitive.*

That is, if $u$ is in the $\rho$-class of $v$, and $v$ is in the $\rho$-class of $z$, $u$ is in the $\rho$-class of $z$.

THEOREM 2.3. *The subset of elements of the group $G(\rho, \omega)$ with no rank or with a rank $\rho < \omega$ form a group $H_\omega$.*

Let $g$ be a subgroup of $G$. Let $\alpha$ be an enumerating index in a simply ordered set. For each $\alpha$ let $h(\alpha)$ be a subgroup of $g$. The group $g$ is said to be a *direct sum*

$$g = \sum_\alpha h(\alpha),$$

if the groups $h(\alpha)$ generate $g$ and if there exists no relation of the form

$$u(\alpha_1) + \cdots + u(\alpha_m) = 0,$$

in which the $\alpha_i$'s are distinct and $u(\alpha_i)$ is a non-null element from the group $h(\alpha_i)$.

The following three theorems depend only on rank conditions I, II, III.

THEOREM 2.4. *If $g$ is a $\rho$-subgroup of $G$ whose ranks form a discrete ascending sequence, $g$ is a direct sum,*

$$g = \sum_\rho g(\rho), \tag{2.1}$$

*of arbitrary maximal subgroups $g(\rho)$ of $g$ with the respective ranks $\rho$ of elements of $g$.*

Let $\rho_1 < \cdots < \rho_n$ be a subset of ranks of elements of $g$, and let $u_1, \cdots, u_n$ be non-null elements of the respective groups $g(\rho_1), \cdots, g(\rho_n)$. It follows from II and III that the sum

$$v = u_1 + \cdots + u_n$$

has the rank $\rho_n$. Hence $v \neq 0$. The groups $g(\rho)$ thus have a $\rho$-subgroup $g'$ of $g$ as a direct sum. It remains to show that $g' = g$.

To that end let $u$ be an arbitrary non-null element of $g$. The theorem will follow if assumption $(\alpha)$ leads to a contradiction.

$(\alpha)$. *Whatever the element $v$ in $g'$, $u - v \neq 0$.*

Set $\rho(u) = r$. Since $g(r)$ is a maximal group of elements of $g$ with rank $r$ it follows from Lemma 2.2 that there exists an element $v_1$ in $g(r)$ such that $u - v_1$ is an element $u_1$ not in $g(r)$. The element $u_1$ has a rank by virtue of $(\alpha)$, and it follows from II that $\rho(u_1) < r$. We now treat $u_1$ as we treated $u$. It follows from $(\alpha)$ that $(\alpha)$ is satisfied if $u_1$ replaces $u$. Proceeding inductively we infer the existence of an infinite sequence of elements $u_1, u_2, \cdots$ with decreasing ranks. This is contrary to the assumption that the ranks of elements in $g$ form a discrete ascending sequence.

Hence $(\alpha)$ is false, $g = g'$, and the theorem is true.

THEOREM 2.5. *If the ranks of elements of $G$ form a discrete ascending sequence, arbitrary maximal groups $g(\rho)$ of elements with the respective ranks $\rho$ have as a direct sum a maximal $\rho$-subgroup of $G$.*

The groups $g(\rho)$ clearly have a $\rho$-subgroup $g$ of $G$ as a direct sum. It remains to show that $g$ is maximal. Let $g^*$ be any maximal $\rho$-subgroup of $G$ containing $g$. Then $g(\rho)$ is a maximal subgroup of elements of $g^*$ with rank $\rho$. It follows from the preceding theorem that $g^*$ is the direct sum of the groups $g(\rho)$, and the proof is complete.

THEOREM 2.6. *Suppose the ranks $\rho$ of $G$ form a discrete ascending sequence, and that $g$ is a maximal $\rho$-subgroup of $G$. If $g(\rho)$ is a subgroup of elements of $g$ with rank $\rho$, chosen so as to be maximal in $g$, then $g(\rho)$ is also maximal in $G$.*

The group $g$ is the sum

$$g = \sum_{\rho} g(\rho) \tag{2.2}$$

of the respective groups $g(\rho)$. If one of these groups, say $g(r)$, is not maximal in $G$ there will exist a subgroup $g'(r)$ of elements of $G$ with rank $r$ such that $g'(r)$ contains $g(r)$ and is maximal in $G$. In (2.2) let $g(r)$ be replaced by $g'(r)$. Let the resulting sum (2.2) then be denoted by $g^*$. It is clear that $g^*$ contains $g$ and is a $\rho$-subgroup of $G$. But $g^*$ is not identical with $g$ because $g'(r)$ is a maximal subgroup of elements of $g^*$ with rank $r$. Hence $g$ is not a maximal $\rho$-subgroup of $G$. From this contradiction we infer that $g(r)$ is maximal in $G$.

Let $g$ be a $\rho$-subgroup of $G$. A base $B$ for $g$ will be termed *$\rho$-proper* if each finite linear combination,[6]

$$v = \delta_i u_i\,, \qquad \delta_i \neq 0, \qquad\qquad (i = 1, \cdots, r)$$

of elements of $B$ satisfies the condition

$$\rho(v) = \max\,[\rho(u_1), \cdots, \rho(u_r)].$$

In general a minimal base for $g$ will not be $\rho$-proper, and $g$ may even admit no such base.

LEMMA 2.3. *A $\rho$-subgroup $g$ of $G$ whose ranks form a discrete ascending sequence admits a $\rho$-proper base.*

For $g$ is a direct sum,

$$g = \sum_{\rho} g(\rho),$$

of the maximal subgroups of elements of $g$ with the respective ranks $\rho$. If $B(\rho)$ is a base for $g(\rho)$, it is clear that the ensemble of the bases $B(\rho)$ is a $\rho$-proper base for $g$. The proof of the lemma is complete.

Let $(u_1\,, u_2\,, \cdots)$ be a finite or infinite set of independent elements of $G$. Let $(v_1\,, v_2\,, \cdots)$ be a second such set with $u_i$ corresponding to $v_i$. Let $g(u)$ and $g(v)$ be the operator groups generated respectively by the elements $u_i$ and $v_i$. If the finite proper linear combination $\delta_i u_i$ of $g(u)$ be made to correspond to $\delta_i v_i$ in $g(v)$, $g(u)$ will thereby be isomorphically mapped on $g(v)$. We shall say that the mapping of the set $(u)$ onto the set $(v)$ has been *isomorphically extended* to the groups $g(u)$ and $g(v)$.

THEOREM 2.7. *Let $(u)$ be the set of elements in a $\rho$-proper base of a group $g(u)$. Let each element $u_i$ of $(u)$ be mapped on an element $v_i$ in the same rank class. The elements $v_i$ are independent and generate a group $g(v)$. The isomorphic extension of the mapping of the base $(u)$ on the base $(v)$ defines a $\rho$-isomorphism between $g(u)$ and $g(v)$.*

[6] We adopt the summation convention of tensor analysis.

The mapping of $g(u)$ onto $g(v)$ is a homeomorphism. Under this homeomorphism a linear combination $x = \delta_i x_i$ $(i = 1, \cdots, r)$ of the elements of $(u)$ corresponds to the element $y = \delta_i y_i$ of $g(v)$ where $x_i$ and $y_i$ are in the same rank class. If $\delta_i = 0$ for no $i$, and if we set

$$\omega = \max [\rho(x_1), \cdots, \rho(x_r)],$$

then $\rho(x) = \omega$ since $(u)$ is a $\rho$-proper base. The element $x - y$ is in the subgroup $H_\omega$ of Theorem 2.3 so that $x - y$ has no rank or a rank $< \omega$. Hence $x$ and $y$ are in the same rank class and $y$ is not null. The homeomorphic mapping of $g(u)$ onto $g(v)$ is accordingly an isomorphism, and since corresponding elements $x$ and $y$ are in the same rank class this isomorphism is a $\rho$-isomorphism.

THEOREM 2.8. *If the ranks $\rho$ of elements of $G$ form a discrete ascending sequence, any two maximal $\rho$-subgroups of $G$ are $\rho$-isomorphic.*

Let $g$ and $g'$ be the two maximal groups. If $g(\rho)$ and $g'(\rho)$ are respectively maximal subgroups of $g$ and $g'$ with the rank $\rho$, then

$$g = \sum_\rho g(\rho), \qquad g' = \sum_\rho g'(\rho),$$

as we have seen in Theorem 2.4. It follows from Theorem 2.6 that $g(\rho)$ and $g'(\rho)$ are maximal groups of elements of $G$ with rank $\rho$, and we can conclude from Theorem 2.1 that $g(\rho)$ and $g'(\rho)$ admit a $\rho$-isomorphism $I(\rho)$. Let $B(\rho)$ be a base for $g(\rho)$ and let $B'(\rho)$ be the base for $g'(\rho)$ which arises by virtue of the preceding $\rho$-isomorphism $I(\rho)$. The logical sums

$$B = \sum_\rho B(\rho), \qquad B' = \sum_\rho B'(\rho)$$

form $\rho$-proper bases for $g$ and $g'$. The isomorphisms $I(\rho)$ map $B$ on $B'$. If this mapping be isomorphically extended $g$ will be mapped $\rho$-isomorphically on $g'$ in accordance with the preceding theorem.

The proof of the theorem is complete.

The following lemma will be useful.

LEMMA 2.4. *If $g$ is an operator subgroup of $G$ and $g_1$ an operator subgroup of $g$, then $g$ is represented by the direct sum*

$$g = g_1 + g_2,$$

*where $g_2$ is a maximal group of elements of $g$, not in $g_1$.*

The groups $g_1$ and $g_2$ have a direct sum. For if $u_1$ and $u_2$ are respectively non-null elements in $g_1$ and $g_2$ and $u_1 + u_2 = 0$, $u_2$ would be in $g_1$.

Moreover $g = g_1 + g_2$. For if $u$ is an arbitrary element in $g$ but not in $g_1$ or $g_2$, there exists an element $v$ in $g_2$ such that $u - v$ is in $g_1$. Thus $u$ is the sum of an element $u - v$ in $g_1$ and an element $v$ in $g_2$.

The proof of the lemma is complete.

THEOREM 2.9. *If $g$ is an operator subgroup of $G$, the rank function $\theta(u)$ defined by $\rho(u)$ as $u$ ranges over $g$ satisfies the rank conditions $R$.*

That $\theta(u)$ satisfies rank conditions I, II, and III is an immediate consequence

of the fact that $\rho(u)$ satisfies $R$, and that $g$ is an operator subgroup of $g$. Condition IV stated for $\theta(u)$ will be denoted by IV$g$.

Condition IV$g$ concerns the group $g(\theta, \omega)$ generated by elements of $g$ for which $\theta(u) \leqq \omega$. Condition IV concerns the group $G(\rho, \omega)$ generated by elements of $G$ for which $\rho(u) \leqq \omega$. Observe that $g(\theta, \omega)$ is a subgroup of $G(\rho, \omega)$. If $u$ and $v$ are elements in $g(\theta, \omega)$ without rank $\theta$, condition IV$g$ affirms that $\theta(u + v) \neq \omega$. We shall see that this true. For $u$ and $v$ are elements in $G(\rho, \omega)$ without rank $\rho$, and it follows from IV that $\rho(u + v) \neq \omega$. Hence $\theta(u + v) \neq \omega$, and the proof of IV$g$ is complete.

### 3. Bi-lexical ranks

Let $s$ and $t$ be real variables which range on the intervals $0 \leqq s \leqq \infty$, $0 \leqq t \leqq \infty$, with $s > t$. Let the pairs $(s, t)$ be ordered lexicographically. That is let $(s, t) = (s', t')$ if and only if $s = s'$, $t = t'$, and

$$(s, t) < (s', t'), \qquad (s', t') > (s, t),$$

if $s < s'$ or if $s = s'$ and $t < t'$. In this section we suppose that for certain non-null elements $u$ of $G$ rank functions $s(u)$ and $t(u)$ are defined. We understand that $t(u)$ exists if and only if $s(u)$ exists. We set

$$r(u) = [s(u), t(u)],$$

and prove the following theorem.

THEOREM 3.1. *Necessary and sufficient conditions that the bi-lexical ranks* $r(u) = [s(u), t(u)]$ *satisfy rank conditions* I, II, III *are as follows*:

(a). *The ranks* $s(u)$ *satisfy conditions* I, II, III.

(b). *For each fixed* $s$ *the rank function* $t^s(u)$ *defined by* $t(u)$ *when* $s(u) = s$ *shall satisfy conditions* I, II, III.

(c). *When* $s(v) < s(u)$, $t(u + v) = t(u)$.

We shall first assume that the ranks $r(u)$ satisfy I, II, III, and show that (a), (b), and (c) are satisfied.

That (a) is satisfied follows directly from the fact that $r(u)$ satisfies I, II, III, and that the pairs $(s, t)$ are ordered lexicographically. The verification of (b) is similar. Finally under the hypothesis of (c),

$$r(u + v) = r(u),$$

since $r(u)$ satisfies II and III. Hence $t(u + v) = t(u)$.

Conversely we shall show that conditions (a), (b), and (c) imply that the ranks $r(u)$ satisfy conditions I to III.

I. It follows from (a) that for $\delta \neq 0$, $s(\delta u) = s(u)$. Setting $s = s(u)$ in (b) we infer from (b) that $t(\delta u) = t(u)$. Hence

$$r(\delta u) = r(u),$$

and I is satisfied by ranks $r(u)$.

II. To establish II for ranks $r(u)$ we may assume without loss of generality

that $r(v) \leqq r(u)$. We must then prove that $r(u + v) \leqq r(u)$. By virtue of (a) we have two cases:

($\alpha$). $s(u + v) < s(u)$,

($\beta$). $s(u + v) = s(u)$.

In case ($\alpha$), $r(u + v) < r(u)$ by virtue of the ordering of the pairs $(s, t)$. In case ($\beta$) we distinguish two subcases:

($\beta'$). $s(v) < s(u)$,

($\beta''$). $s(u) = s(v)$.

In case ($\beta'$), $t(u + v) = t(u)$ in accordance with (c) so that $r(u + v) = r(u)$, and II holds. In case ($\beta''$), $t(u + v) \leqq t(u)$ by virtue of (b). Hence

$$r(u + v) \leqq r(u).$$

Thus II holds in all cases.

III. To establish III we must show that $r(u + v)$ exists, provided $r(u)$ and $r(v)$ exist and are unequal. Without loss of generality we can assume that $r(u) > r(v)$. We consider two cases:

(m). $s(u) > s(v)$,

(n). $s(u) = s(v)$, $t(u) > t(v)$.

In case ($m$), it follows from (a) that $s(u + v)$ exists. Hence $r(u + v)$ exists. In case (n), $t(u + v)$ exists by virtue of (b). Hence $r(u + v)$ exists in both cases, and the proof of III is complete.

The ranks $t(u)$ need not satisfy conditions I to III even when the ranks $r(u)$ satisfy these conditions. This is shown by the following example.

EXAMPLE 3.1. Let $G$ consist of the elements

$$0, \quad u, \quad v, \quad u + v,$$

the field $\Delta$ being the integers mod 2. Suppose that

$$r(u) = (3, 0), \qquad r(v) = (3, 0), \qquad r(u + v) = (2, 1).$$

The ranks $r$ satisfy I, II, and III. However

$$t(u) = 0, \qquad t(v) = 0, \qquad t(u + v) = 1$$

so that

$$t(u + v) > \max [t(u), t(v)].$$

Thus the ranks $t$ do not satisfy II.

*Spans* $s - t$. An element $u$ with rank $r = (s, t)$ will be said to have the span $s - t$, understanding that $s - t$ is infinite if $s$ is infinite. It is clear that $s - t$ is constant as $u$ varies over any $r$-class.

## 4. Invariant classes

Let $A$ be the class of all elements of $G$ with ranks in the set $[\rho]$. A subset $E$ of $A$ will be termed *$\rho$-invariant* if whenever $u$ is in $E$ every element in the $\rho$-class of $u$ is in $E$. Let $E(\rho, \omega)$ denote the group generated by elements of $E$ with

$\rho \leqq \omega$. We shall be concerned with $\rho$-invariant subsets $E$ which satisfy the following condition:

V. *If $u$ and $v$ are in $E(\rho, \omega)$, if*

$$\rho(u) = \rho(v) = \rho(u + v) = \omega,$$

*and $u$ and $v$ are not in $E$, then $u + v$ is not in $E$.*

EXAMPLE 4.1. Let $A$ be a set of elements of $G$ with bi-lexical ranks $(s, t)$ satisfying rank conditions $R$. The subset of elements of $A$ with spans $s - t > e$ forms an invariant subset $E$ of $A$ satisfying V.

THEOREM 4.1. *Let $E$ be an invariant subset of $A$ which satisfies* V. *The rank function $\eta(u)$ defined by $\rho(u)$ when $u$ is in $E$ satisfies the rank conditions $R$.*

That $\eta(u)$ satisfies conditions I and II of $R$ follows immediately from the fact that the rank function $\rho(u)$ satisfies I and II, and that $E$ is an operator subset of $A$. In III, $\eta(u) \neq \eta(v)$ by hypothesis. Hence $\rho(u) \neq \rho(v)$. Suppose $\rho(u) > \rho(v)$. Then $\rho(u + v)$ exists, and $u + v$ and $u$ are in the same $\rho$-class. Since $E$ is a $\rho$-invariant subclass of $A$, $u + v$ must be in $E$ because $u$ is in $E$. Hence $\eta(u + v)$ exists, and III is satisfied.

Let IV$\rho$ and IV$\eta$ denote the conditions IV stated respectively for ranks $\eta$ and ranks $\rho$. Note that $G(\eta, \omega)$ in IV$\eta$ is a subgroup of $G(\rho, \omega)$ in IV$\rho$. We seek to establish IV$\eta$.

Condition IV$\eta$ may be put in the following form:

($\alpha$). *If $u$ and $v$ are in $G(\eta, \omega)$ and $\eta(u + v) = \omega$, at least one of the elements $u$ and $v$ is in $E$.*

The elements $u$ and $v$ are in $G(\rho, \omega)$ and $\rho(u + v) = \omega$. It follows from IV$\rho$ that not both $\rho(u)$ and $\rho(v)$ can fail to exist. We consider two cases:

$$\text{CASE a. } \rho(u) < \omega, \qquad \text{CASE b. } \rho(u) = \omega,$$

the remaining cases arising upon replacing $u$ by $v$.

CASE a. Here $\rho(u) < \omega$, $\rho(u + v) = \omega$ so that $v$ and $u + v$ are in the same $\rho$-class. But $u + v$ is in $E$ since $\eta(u + v)$ exists in ($\alpha$), and $E$ is $\rho$-invariant. Hence $v$ is in $E$.

CASE b. Here $\rho(u) = \omega$. If $\rho(v)$ fails to exist or $\rho(v) < \omega$, then $u + v$ and $u$ are in the same $\rho$-class and $u$ is in $E$ as in Case a. There remains the case

$$\rho(u) = \rho(v) = \eta(u + v) = \omega.$$

It follows from V that at least one of the elements $u$ and $v$ is in $E$.

Thus ($\alpha$) is true in all cases and the proof of IV$\eta$ is complete. This completes the proof of the theorem.

Let $b$ be an operator subgroup of $G$ and $c$ be a maximal group of elements of $G$ not in $g$. Then $b + c = G$ as is well-known. Theorem 4.2 which follows is a theorem of this type, but of a less regular form. The reader may find it simpler to postpone the study of Theorem 4.2 until its application in the proof of Theorem 11.3. The appropriateness of the form of Theorem 4.2 will then be clear. It has been included in this section to bring out the fact that it is independent of the topological aspects of this paper.

Let $A$ be the class of all elements of $G$ with rank, and suppose that $A$ is the sum of two disjoint invariant subclasses $B$ and $C$ with the following properties:

(i). The sum of two elements in $B$ in different $\rho$-classes is in $B$.

(ii). The sum of two elements in $C$ with different ranks $\rho$ is in $C$.

(iii). The sum of an element in $B$ and an element in $C$ is in $A$.

THEOREM 4.2. *Let $B$ and $C$ be invariant subclasses of $A$ satisfying the above conditions, and let $b$ and $c$ be respectively maximal groups of elements in $B$ and $C$. If the ranks of $C$ form a discrete ascending sequence, $b + c$ is a maximal group of elements of $A$.*

That $b + c$ is a group of elements in $A$ follows from (iii). It remains to show that $b + c$ is maximal in $A$. To that end let $u$ be an arbitrary element in $A$, and assume that the following statement is true.

($\alpha$). *If $v$ is an arbitrary element in $b + c$, $u - v$ is in $A$.*

We shall show that ($\alpha$) leads to a contradiction.

If $u$ is in $C$, there exists an element $x$ in $c$ such that $u - x$ is not in $C$, since $c$ is maximal. Setting $v_1 = u - x$ we see that $v_1$ is in $B$ by virtue of ($\alpha$). The elements $u$ and $x$ are both in $C$, but $v_1$ is not in $C$. It follows from (ii) that $\rho(x) = \rho(u)$. But $v_1 = u - x$, and it follows from rank condition II that

$$\rho(v_1) \leqq \rho(u) \qquad (v_1 \subset B). \tag{4.1}$$

In case $u$ is in $B$ we set $v_1 = u$ and $x = 0$, so that (4.1) holds for any element $u$ in $A$.

Since $v_1$ is in $B$ and $b$ is maximal there exists an element $y$ in $b$ such that $v_1 - y$ is not in $B$. Set

$$u_1 = v_1 - y = u - x - y.$$

It follows from ($\alpha$) that $u_1$ is in $C$. Since $v_1$ and $y$ are in $B$ but $v_1 - y$ is not in $B$, it follows from (i) that $v_1$ and $y$ are in the same $\rho$-class. But $u_1 = v_1 - y$ and it follows from the definition of a $\rho$-class that $\rho(u_1) < \rho(v_1)$. Hence

$$\rho(u_1) < \rho(u).$$

Observe that ($\alpha$) holds if $u$ is replaced by $u_1$. Proceeding inductively we infer the existence of an infinite sequence of elements $u_n$ in $C$ with ranks $\rho(u_n)$ which decrease with $n$. This is impossible since the ranks of $C$ form a discrete ascending sequence.

We conclude that ($\alpha$) is false and that $b + c$ is maximal in $A$. The proof of the theorem is complete.

## 5. Vietoris chains and cycles[7]

The space $M$ shall be a metric space of elements $p, q, r, \cdots$ with a distance function $pq$ satisfying the usual conditions:

I. $pp = 0$, II. $pq \neq 0$ if $p \neq q$, III. $pr \leqq pq + rq$.

[7] Vietoris, L., *Über den höheren Zusammenhang kompakter Räume und eine Klasse von Zusammenhangstreuen Abbildungen*, Mathematische Annalen 97 (1927), pp. 454–472. Further references can be found in M1.

We shall be concerned with formal $k$-chains and $k$-cycles (Vietoris) defined as in M1, §1. We take over the definitions of pages 4–7 including definitions of the terms algebraic $k$-chains, algebraic $k$-cycles mod $B$ on $C$, (Vietoris) $k$-cycles mod $B$ on $C$, homologies $u \sim 0$ and $u \sim v$ mod $B$ on $C$, carriers of cycles and homologies, homology classes, the boundary $\beta u$ of an algebraic $k$-chain $u$ or a formal $k$-chain $u$. We combine $k$-chains and $k$-cycles linearly with coefficients $\delta$ in an arbitrary field $\Delta$.

Let $B$ and $C$ be compact sets on $M$ with $B \subset C$. Let $B_e$ and $C_e$ be respectively $e$-neighborhoods of $B$ and $C$. Let $u = (u_n)$ be a $k$-cycle mod $B$ on $C$ with algebraic component $k$-cycles $u_n$ mod $B$ on $C$. We shall make use of the following well-known lemma.

LEMMA 5.1. *If $u = (u_n)$ is a $k$-cycle* mod $B$ *on $C$ such that $u \sim 0$* mod $B_e$ *on $C_e$ for each positive $e$, then $u \sim 0$* mod $B$ *on $C$.*

Corresponding to an arbitrary positive constant $e$ there exists by hypothesis an integer $N$ so large that for $n > N$,

$$u_n \sim_e 0 \qquad (\text{mod } B_e \text{ on } C_e).$$

For such values of $n$ we infer the existence of an algebraic $(k+1)$-chain $w_n$ on $C_e$ of norm $e$ and an algebraic $k$-chain $v_n$ on $B_e$ of norm $e$ such that

$$\beta w_n = u_n + v_n .$$

From this relation we see that $\beta u_n + \beta v_n = 0$, so that $\beta v_n$ is on $B$. Let $fv_n$ be the algebraic chain obtained from $v_n$ by replacing each vertex $x$ of $v_n$ not on $B$ by a nearest point $fx$ on $B$. Let $Dv_n$ be the deformation chain of M1, page 9, corresponding to this mapping $fx$. As in M1, (1.3),

$$\beta D v_n = v_n - fv_n - D\beta v_n .$$

But $\beta v_n$ is on $B$ so that $D\beta v_n = 0$. If we set $w_n' = w_n - Dv_n$ we find that

$$\beta w_n' = u_n + fv_n ,$$

where $w_n'$ has the norm $3e$, and $fv_n$ is on $B$.

We now replace each vertex of $w_n'$ on $C_e$ but not on $C$ by a nearest vertex on $C$, obtaining thereby a $(k+1)$-chain $y_n$ on $C$ of norm at most $5e$. The chain $u_n + fv_n$ is on $C$ and is unchanged. Hence

$$\beta y_n = u_n + fv_n ,$$

and since $fv_n$ is on $B$, $u \sim 0$ mod $B$ on $C$.

Let $A$, $B$, and $C$ be three compact sets of $M$ with $A \subset B \subset C$. We shall make use of the three following theorems. Proofs can be found in M2, page 396, making use of two fundamental lemmas due to E. Čech.

THEOREM 5.1. *If $u$ is a $k$-cycle* mod $A$ *on $C$, homologous to* 0 mod $B$ *on $C$, then the homology class of $u$* mod $A$ *on $C$ contains a $k$-cycle* mod $A$ *on $B$.*

THEOREM 5.2. *If a $k$-cycle $u$* mod $A$ *on $B$ is homologous to* 0 mod $A$ *on $C$, there exists a $(k+1)$-cycle* mod $B$ *on $C$ such that $\beta v - u$ is on $A$.*

THEOREM 5.3. *If $u$ is a $k$-cycle* mod $A$ *on* $C$ *and* $\beta u \sim 0$ *on* $A$, *there exists an absolute $k$-cycle $v$ on $C$ such that* $u \sim v$ mod $A$ *on* $C$.

The dimension of a maximal group of $n$-cycles on $M$, non-bounding on $M$ will be termed the $n^{\text{th}}$ *connectivity* of $M$. If $A \subset B \subset M$, the dimension of a maximal group of $n$-cycles on $A$ non-bounding on $B$ will be termed the $n^{\text{th}}$ *connectivity of $A$ on $B$*. We shall be concerned with cases where this dimension is finite and obviously uniquely defined.

## 6. The function $F$

The function $F(p)$ shall be single-valued on $M$ and such that $0 \leqq F < 1$. A function $J$ on $M$ such that $0 \leqq J < \infty$ can be reduced to a function $F$ by a transformation

$$F = \frac{J}{1 + J}.$$

Let $c$ be a finite constant. The subsets of $M$ on which $F \leqq c$ will be denoted by $F_c$. A set $A$ will be said to be *definitely on* $F_c$ (written d-on $F_c$) if $A$ is on $F_{c-e}$ for some positive $e$. The phrase d-mod $F_c$ shall mean mod $F_{c-e}$ for some positive $e$. A set d-on $F_c$ will be said to be *d-below* $c$.

The space $M$ and the function $F$ shall satisfy three principal hypotheses of which the first is as follows.

*Bounded compactness. For each constant $c < 1$, the sets $F_c$ shall be compact.*

This postulate implies that the function $F$ is lower semi-continuous. It is always fulfilled when $M$ is compact and $F$ is lower semi-continuous.

Let $p$ be a point of $M$ at which $F(p) = c$. The set $M$ will be said to be *locally $F$-connected* of order $r$ at $p$ if corresponding to each positive constant $e$ there exists a positive constant $\delta$ such that each singular $r$-sphere on the $\delta$-neighborhood of $p$ and on $F_{c+\delta}$ bounds an $(r+1)$-cell of norm $e$ on $F_{c+e}$. It might be thought that local $F$-connectedness of order $r$ is the same as local connectedness of order $r$ in the space of the pairs $[p, F(p)]$. That this is not the case is shown by the following example.

EXAMPLE 6.1. Let the space $M$ be the semi-circular disc $x^2 + y^2 \leqq 1$, $y \geqq 0$. Suppose that $F(x, y) = y$ except when $x = 0$. Let $F(0, y) \equiv 0$. The space $M$ is locally $F$-connected of all orders, but the set of points $[x, y, F(x, y)]$ regarded as a set in euclidean three-space $[x, y, z]$ is not locally 0-connected at the points $[0, y, 0]$ at which $y > 0$.

*Local $F$-connectedness. We assume that $M$ is locally $F$-connected of all orders at each point $p$ of $M$.*

If $M$ is locally connected of order $r$ and if $F$ is continuous, $M$ is locally $F$-connected of order $r$.

The whole space $M$ is not in general compact. That each homology class contains a cycle on a compact subset $F_c$ of $M$ is part of the implication of the following hypotheses.

*$F$-regularity. Let $a$ and $b$ be constants with $a < b < 1$. In each non-trivial*

*homology class on $M$* mod $F_b$ *there shall exist a cycle* mod $F_b$ *on some compact subset $F_c$ of $M$ where $c < 1$. Moreover each relative $k$-cycle $v$* mod $F_a$ *on $F_b$ which satisfies an homology*

$$v \sim 0 \qquad (\text{mod } F_b) \tag{6.1}$$

*on $M$ shall satisfy a similar homology on some compact subset $F_c$ of $M$ where $c < 1$.*

A simple example in which this postulate is not satisfied follows.

EXAMPLE 6.2. Let $M$ be the circle

$$x^2 + (y - 1)^2 = 1. \tag{6.2}$$

We see that $0 \leqq y \leqq 2$. Let $F = y/2$ except at the point $(x, y) = (0, 2)$. Set $F(0, 2) = 0$. The circle is a carrier of a non-bounding 1-cycle which is homologous to no cycle with carrier on which $F$ is bounded from 1.

These three postulates are satisfied for all ordinary variational problems. See M1, Part III. They are also satisfied by a suitably chosen space and function $F$ in minimal surface theory. See Morse and Tompkins, loc. cit.

Let $a, b, c$ be three constants such that

$$a < b < c < 1.$$

We shall prove two theorems which would be false if stated for singular cycles, and which embody the principle of the "accessibility" of the sets $F_c$ with respect to homologies.

THEOREM 6.1. *If a $k$-cycle $u$* mod $F_a$ *on $F_c$ is such that*

$$u \sim 0 \qquad (\text{on } F_c \text{ mod } F_{b+e}) \tag{6.3}$$

*for every positive $e$, the homology class of $u$* mod $F_a$ *on $F_c$ contains a $k$-cycle* mod $F_a$ *on $F_b$.*

It follows from Lemma 5.1 and (6.3) that

$$u \sim 0 \qquad (\text{on } F_c \text{ mod } F_b).$$

We now turn to Theorem 5.1 and set

$$A = F_a, \qquad B = F_b, \qquad C = F_c, \tag{6.4}$$

noting that these sets are compact. Theorem 6.1 follows from Theorem 5.1.

If $a < 0$, u becomes an absolute cycle and there is a cycle in its homology class on $F_b$.

THEOREM 6.2. *If a $k$-cycle $u$* mod $F_a$ *on $F_b$ is homologous to* 0 mod $F_a$ *on $F_{c+e}$ for every positive $e$, there exists a $(k + 1)$-cycle $v$* mod $F_b$ *on $F_c$ such that $\beta v - u$ is on $F_a$.*

It follows from Lemma 5.1 that $u \sim 0$ mod $F_a$ on $F_c$. We now apply Theorem 5.2, setting $A, B, C$ equal to $F_a, F_b, F_c$ respectively. Theorem 6.2 results.

If $a < 0$, $u$ is an absolute cycle and there exists a $(k + 1)$-cycle $v$ mod $F_b$ on $F_c$ such that $\beta v = u$.

THEOREM 6.3. *Let $a$ and $c$ be positive constants such that $a < c < 1$. The $k^{\text{th}}$ connectivity $R^k(a, c)$ of $F_a$ on $F_c$ is finite.*

This is a consequence of the hypotheses of bounded compactness and of local $F$-connectedness, as follows from Theorem 6.1 of M1, page 28.

## 7. $F$-homology classes and ranks $r$

An absolute $k$-cycle $u$ which lies definitely on some subset $F_c$ of $M$ for which $c < 1$ and satisfies the condition

$$u \nsim 0 \qquad \text{(d-on } F_c\text{)} \tag{7.1}$$

will be termed an *F-cycle*, and will be said to possess the *homology bound* $c$. The least upper bound of the homology bounds of $u$ will be called the *superior cycle limit*[8] $s(u)$ *of* $u$. In case $u$ is non-bounding we understand that $s(u)$ is infinite. In case $u$ is bounding it follows from the condition of $F$-regularity that $s(u) < 1$. Thus $s(u)$ never equals 1. It follows from the definitions involved that $s(u)$ is an homology bound of $u$ in case $u$ is bounding. Given an $F$-cycle $u$, the greatest lower bound of numbers $b < s(u)$ such that

$$u \sim 0 \qquad \text{(d-on } F_s \text{ mod } F_b\text{)} \tag{7.2}$$

will be called the *inferior cycle limit* $t(u)$ of $u$.

If $s(u)$ is regarded as a rank, two $F$-$k$-cycles $u$ and $v$ are in the same $s$-class if and only if for $s = s(u)$

$$u \sim v \qquad \text{(d-on } F_s\text{).} \tag{7.3}$$

An $s$-class becomes an homology class when $s = \infty$. An $s$-class will accordingly be termed an *F-homology* class.

If $u$ is an $F$-cycle the pair $[s(u), t(u)]$ will be regarded as a rank $r(u)$ of $u$. We order these ranks lexicographically. If two elements $u$ and $v$ are in the same $F$-homology class they are in the same $r$-class but not conversely. A necessary and sufficient condition that elements $u$ and $v$ with the same rank $r = (s, t)$ be in the same $r$-class is that

$$u \sim v \qquad \text{(d-on } F_s \text{ d-mod } F_t\text{).} \tag{7.4}$$

It is seen that (7.3) implies (7.4), but not conversely.

Observe that null cycles have no $s$-rank or $t$-rank.

We continue with the following lemma.

LEMMA 7.1. *The ranks* $s(u)$ *of* $k$*-cycles satisfy rank conditions* $R$.

The underlying group $G$ is here the group of all absolute $k$-cycles.

I. To show that $s(\delta u) = s(u)$ for $\delta \neq 0$ we must show that the condition $u \sim 0$ d-on $F_c$ is equivalent to the condition $\delta u \sim 0$ d-on $F_c$. The first condition clearly implies the second. But the second condition implies the first. For any coefficient in the field $\Delta$ can be divided by a non-null $\delta$. Thus I holds.

II. In II, it is assumed that $u$, $v$, and $u + v$ have ranks $s$. Let $c$ be the maximum of $s(u)$ and $s(v)$. By virtue of the definition of superior cycle limits

$$u \sim 0, \qquad v \sim 0, \qquad \text{(on } F_{c+\epsilon}\text{)}$$

[8] In M1, $s(u)$ was inadvertently used in another sense. We revert to the usage of M2.

for each positive constant $e$, so that

$$u + v \sim 0 \qquad (\text{on } F_{c+e}).$$

Hence $s(u + v) \leqq c$, and II is satisfied.

III. We suppose that $s(u) > s(v)$ and let $c$ be any homology bound of $u$ between $s(u)$ and $s(v)$. Then $v \sim 0$ and $u \nsim 0$ d-on $F_c$ so that $u + v \nsim 0$ d-on $F_c$. Thus $c$ is an homology bound of $u + v$. Hence $s(u + v)$ exists.

IV. In the notation of IV let $u$ and $v$ be elements in $G(s, \omega)$. The cycles $u$ and $v$ are d-on $F_b$ where $b$ is some constant $< \omega$. In IV, $b$ is an homology bound of neither $u$ nor $v$, and so it cannot be an homology bound of $u + v$. Hence $s(u + v) < b < \omega$ if $s(u + v)$ exists. We infer the truth of IV.

LEMMA 7.2. *Let $c$ be a value assumed by the rank function $s$. The rank function $t^c(u)$ defined by $t(u)$ for those $k$-cycles for which $s = c$ satisfies rank conditions* I, II, III.

I. Condition I follows from the fact that for $\delta \neq 0$ relation (7.2) holds if and only if it holds when $u$ is replaced by $\delta u$.

II. Let $b$ be any constant such that

$$c > b > \max [t(u), t(v)] \qquad (s(u) = s(v) = c). \tag{7.5}$$

Then

$$u \sim 0, \qquad v \sim 0, \qquad (\text{d-on } F_c \bmod F_b)$$

in accordance with (7.2). Hence $u + v$ satisfies a similar homology. We conclude that $t(u + v) \leqq b$. But $b$ is any constant for which (7.5) holds so that

$$t(u + v) \leqq \max [t(u), t(v)],$$

and the proof of II is complete.

III. We here assume that $s(u) = s(v) = c$, and that $t(u) \neq t(v)$. We seek to prove that $s(u + v) = c$ so that $t^c(u + v)$ exists.

We suppose that $t(u) < t(v)$. Then for each constant $b$ between $t(u)$ and $t(v)$,

$$u \sim 0, \qquad u + v \sim v \nsim 0, \qquad (\text{d-on } F_c \bmod F_b)$$

in accordance with (7.2). In particular

$$u + v \nsim 0 \qquad (\text{d-on } F_c)$$

so that $s(u + v)$ exists and equals $c$. Hence $t^c(u + v)$ exists, and the proof of III is complete.

THEOREM 7.1. *The ranks $r = (s, t)$ of $F$-$k$-cycles satisfy the rank conditions $R$.*

That ranks $r$ satisfy conditions I, II, III will follow from Theorem 3.1. For conditions (a) and (b) are satisfied in Theorem 3.1 by virtue of the two preceding lemmas. Condition (c) of Theorem 3.1 is that when $s(v) < s(u)$, $t(u) = t(u + v)$. But when $s(v) < s(u)$, $s(u + v) = s(u)$. Setting $a = s(u)$,

$$v \sim 0, \qquad u + v \sim u, \qquad (\text{d-on } F_a).$$

Hence $t(u + v) = t(u)$, and condition (c) is satisfied. The ranks $r$ thus satisfy conditions I, II, III.

To establish IV let $\omega = (\omega_1, \omega_2)$ be a particular rank $r$. The group $G(r, \omega)$ is a subgroup of the group $G(s, \omega_1)$. If $u$ and $v$ are two elements of $G(r, \omega)$, $u$ and $v$ are in $G(s, \omega_1)$. Since ranks $s$ satisfy IV, the hypothesis that $s(u + v) = \omega_1$ implies that $s(u)$ or $s(v)$ exists. Hence $r(u)$ or $r(v)$ exists, and the proof of IV is complete.

An $F$-cycle $u$ with rank $r = (s, t)$ will be termed *canonical* if $u$ is on $F_t$. Let $c$ be an arbitrary constant $c < 1$. A $k$-cycle on $F_c$ will be termed *canonical relative* to $F_c$ if, when $F_c$ is regarded as the whole space $M$, $u$ is canonical. If $\tau$ is the inferior cycle limit of $u$ relative to $F_c$ it is clear that $\tau \geqq t(u)$.

LEMMA 7.3. *Let $H$ be an $F$-homology class of $k$-cycles with rank $r = (s, t)$ and let $c$ be a constant between $s$ and $t$ and $< 1$. Let $\eta$ be an arbitrary positive constant. The class $H$ contains at least one $k$-cycle which is canonical relative to $F_c$, and which possesses an inferior cycle limit $\tau$ relative to $F_c$ such that $\tau < t + \eta$.*

By virtue of the definition of $t$ there exists a constant $b < t + \eta$ and $< c$ such that

$$u \sim 0 \qquad \text{(d-on } F_s \text{ mod } F_b\text{)}.$$

It follows from Theorem 5.1 that the homology class of $u$ d-on $F_s$ contains a $k$-cycle $v$ on $F_b$. Let $\tau$ be the inferior cycle limit of $v$ relative to $F_c$. It is clear that

$$t \leqq \tau \leqq t + \eta.$$

By virtue of the definition of $\tau$,

$$v \sim 0 \qquad \text{(on } F_c \text{ mod } F_{\tau+e}\text{)}$$

for every $e > 0$. It follows from Theorem 6.1 that the homology class of $v$ on $F_c$ contains a $k$-cycle $z$ on $F_\tau$. The cycle $z$ is canonical relative to $F_c$, and the proof of the lemma is complete.

THEOREM 7.2. *Each $F$-homology class of $k$-cycles contains at least one canonical $k$-cycle.*

Let $(s, t)$ be the cycle limits of the given $F$-homology class $H$, and let $c < 1$ be a constant such that $s > c > t$. By virtue of the preceding lemma $H$ contains a sequence of $k$-cycles $z_n$, $n = 1, 2, \cdots$, with the following properties. The cycle $z_n$ is canonical relative to $F_c$ and the inferior cycle limit $\tau_n$ of $z_n$ relative to $F_c$ tends to $t$ as $n$ becomes infinite.

If the theorem is false $\tau_n > t$ for each $n$, and we can suppose without loss of generality that $\tau_n$ decreases with $n$. Let $a$ be a constant such that $t < a < c$. We can suppose that $\tau_n < a$ for each $n$. Hence $z_n$ is on $F_a$. Relative to $F_c$ the bi-lexical rank of $z_n$ is $(\infty, \tau_n)$. Every proper linear combination of the cycles $z_n$ has a rank relative to $F_c$ of the form $(\infty, \tau)$. Thus the cycles $z_n$ are on $F_a$ and independent with respect to bounding on $F_c$. This is impossible by virtue of Theorem 6.3.

We infer the truth of the theorem.

We shall give an example showing that the preceding theorem would not be true were $M$ not locally $F$-connected of the different orders $r$.

EXAMPLE 7.1. Let $M$ be the closure in the space $(x, y)$ of the curve

$$3y = (1 - x) \sin \frac{1}{x} + 1 \qquad (0 < x \leqq 1).$$

On $M$, $0 \leqq y \leqq 2/3$, and $M$ is compact. Let $F = y$ on $M$. The space $M$ is not locally connected of order 0. Let $u$ and $v$ be respectively Vietoris 0-cycles with the points (0, 0), (1, 1/3) as carriers. Let $z$ be the cycle $u - v$. It appears that $s(z) = 2/3$, $t(z) = 0$. But there is no 0-cycle on $F = 0$ homologous to $z$ d-on $F \leqq 2/3$.

As in the general theory of bi-lexical ranks $r = (s, t)$ we term the difference $s - t$ the *span* of an $F$-cycle. Theorem 4.1 thus gives the following. (See Example 4.1).

THEOREM 7.3. *Let $E$ be the subset of $F$-cycles with spans $s - t > e$. The rank function $r_e(u)$ defined by $r(u)$ for $u$ in $E$ satifies the rank conditions $R$.*

We shall continue with the following lemma.

LEMMA 7.4. *The ranks $r$ of bounding $F$-$k$-cycles with spans $> e > 0$ are at most finite in number.*

Since the $F$-cycles $u$ of the lemma are bounding $s(u) < 1$, and since their spans exceed $e$,

$$t(u) < s(u) - e < 1 - e.$$

Without loss of generality we can suppose that the $F$-cycles of the lemma are canonical, since there is a canonical $F$-cycle in each $F$-homology class, and hence a canonical $F$-cycle with each different rank. Let $b$ and $a$ be constants such that

$$0 < b - a < e, \qquad b \leqq 1 - e.$$

Each rank $t(u)$ of a cycle of the lemma lies on an interval of the form

$$a < t < b. \tag{7.6}$$

The set $T$ of such ranks can be covered by a finite set of intervals of this form.

For $a$ and $b$ fixed the number of different ranks $t$ in $T$ and on (7.6) must be finite. For

$$t < b, \qquad s > t + e > a + e,$$

so that the corresponding canonical $F$-cycles lie on $F_b$ and are independent on $F_{a+e}$. The number of such cycles is at most $R(b, a + e)$. Since $a + e > b$ this number is finite in accordance with Theorem 6.3, and the proof of the lemma is complete.

THEOREM 7.4. *The dimension of a maximal group $g$ of bounding $F$-cycles with a given rank $r$ is finite.*

Suppose $r$ has the form $(s, t)$. Let $B$ be a base for $g$. Let each element $v_i$

in $B$ be replaced by a canonical $F$-cycle $u_i$ in the $F$-homology class of $v_i$. We have

$$u_i \sim v_i \qquad \text{(d-below } s\text{)}.$$

The cycles $v_i$ are independent d-below $s$. Hence the cycles $u_i$ are independent d-below $s$. Let $c$ be any constant between $s$ and $t$ and $< 1$. The cycles $u_i$ are on $F_t$ and independent on $F_c$. Their number is at most $R(t, c)$, and the theorem follows from the finiteness of $R(t, c)$.

THEOREM 7.5. *The dimension of a maximal group $g$ of bounding $F$-cycles with spans $> e > 0$ is finite.*

By virtue of Lemma 7.4 the ranks $r$ of elements of $g$ are finite in number. The group $g$ is accordingly a direct sum,

$$g = \sum_r g(r),$$

of maximal subgroups of g with the respective ranks of elements of $g$. The dimension of each such group $g(r)$ is finite in accordance with the preceding lemma. Hence the dimension of $g$ is finite, and the proof of the theorem is complete.

THEOREM 7.6. *In any infinite sequence $(s_n, t_n)$ of distinct ranks of $F$-cycles with spans $> e > 0$, $t_n$ tends to 1 as $n$ becomes infinite.*

It follows from Theorem 7.2 that there exists a canonical $k$-cycle $u_n$ with rank $(s_n, t_n)$. The cycle $u_n$ is on $F \leqq t_n$.

If the theorem were false, the values of $t_n$ would have a cluster value $a < 1$. Let $b$ and $c$ be constants such that

$$a < b < c < a + e \qquad (c < 1).$$

There will exist an infinite subset of the values of $n$ for which $t_n < b$ and $s_n > c$. The corresponding cycles $u_n$ will be on $F_b$ and be independent on $F_c$. This is impossible since the connectivities of $F_a$ on $F_c$ are finite.

The theorem follows.

## 8. $k$-Caps and their cap heights

Let $a$ be a constant at most 1. If $u$ is a $k$-cycle d-mod $F_a$ on $F_a$, an homology of the form

$$u \sim 0 \qquad \text{(on } F_a \text{ d-mod } F_a\text{)} \tag{8.1}$$

will be called an *$a$-homology*. A $k$-cycle on $F_a$ d-mod $F_a$ not $a$-homologous to 0 will be called a $k$-cap with *cap height $a$*. We write $a = a(u)$. We note that $a(u)$ is uniquely determined by the formal $k$-chain $w$ whose components are those of $u$. In fact $a(u)$ is the greatest lower bound $c$ of numbers $b$ such that $w$ is on $F_b$. For it is impossible that $a(u) > c$ since $u$ would then satisfy an $a$-homology. It is equally impossible that $a(u) < c$ since this would imply that $w$ is on $F_a$, contrary to the definition of $c$. That no cap height equals 1 follows at once from the condition of $F$-regularity.

We regard $a(u)$ as a new rank function defined for a subset of elements of the group $G$ of formal chains. It is important that we show that $a(u)$ satisfies the rank conditions $R$, for from this fact it follows that the maximal groups which appear later have dimensions which are independent of their mode of formation.

THEOREM 8.1. *The cap heights $a(u)$ satisfy the rank conditions $R$.*

That I is satisfied follows as for the ranks $s(u)$. That II is satisfied follows from the fact that $a(u)$ is the greatest lower bound of numbers $b$ such that the components of $u$ are on $F_b$. Turning to III, suppose that $a(u) < a(v)$. Then $u$ is $a(v)$-homologous to zero and $v$ is not $a(v)$-homologous to zero. Hence $u + v$ is not $a(v)$-homologous to zero. Since $u + v$ is on $F_a$ for $a = a(v)$, $a(v)$ is a cap height of $u + v$, and III is proved. In IV, $u$ and $v$ are elements of $G(a, \omega)$, not $k$-caps. Hence $u$ and $v$ are $\omega$-homologous to zero. It follows that $u + v$ is $\omega$-homologous to zero, and IV is true.

The proof of the theorem is complete.

LEMMA 8.1. *Let $u$ be a $k$-cycle $d$-mod $F_a$ on $F_a$. If $u$ is $a$-homologous to zero $\beta u$ is homologous to zero $d$-below $a$.*

Under the hypothesis of the lemma, there exists a formal $(k + 1)$-chain $w$ on $F_a$ such that

$$\beta w = u - x, \tag{8.2}$$

where $x$ is a formal $k$-chain d-below $a$. Applying the operator $\beta$ to both sides of (8.2) we find that $\beta u = \beta x$, and the lemma follows directly.

*Cap class.* In accordance with the general definitions of a rank class, two $k$-caps with the same cap height $a$ are said to be in the same $a$-class or *cap class* if $u - v$ satisfies an $a$-homology.

A $k$-cap $u$ with cap height $a$ will be said to be *linkable* if

$$\beta u \sim 0 \qquad \text{(d-below } a\text{)}. \tag{8.3}$$

Two $k$-caps $u$ and $v$ in the same cap class will both be linkable or both non-linkable. For if $u$ and $v$ define the same cap class, $u - v$ is $a$-homologous to zero, and

$$\beta u - \beta v \sim 0 \qquad \text{(d-below } a\text{)}$$

in accordance with the preceding lemma. It follows that (8.3) holds for $u$ if and only if (8.3) holds for $v$.

*Superior cap limit $\sigma(u)$.* Let $u$ be a $k$-cap with cap height $a$. The least upper bound $\sigma$ of numbers $b \geqq a$ such that

$$u \nsim 0 \qquad \text{(on } F_b \text{ d-mod } F_a\text{)} \tag{8.4}$$

will be called the *superior cap limit $\sigma(u)$* of $u$. It is clear that

$$\sigma(u) \geqq a(u).$$

If the numbers $b$ satisfying (8.4) are unbounded, we set $\sigma(u) = \infty$. It appears that two $k$-caps in the same cap class have the same superior cap limit. There exist $k$-caps for which $\sigma(u) = a(u)$, as the following example shows.

EXAMPLE 8.1. Let $M$ be the interval $0 \leqq x \leqq 1$. Let $F(0) = 1/2$ and

$$F(x) = \frac{x}{2} \sin \frac{1}{x} + \frac{1}{2} \qquad (0 < x \leqq 1).$$

The Vietoris 0-cycle $u$ with carrier $x = 0$ is such that $\sigma(u) = a(u) = F(0)$.

No $k$-cap $u$ has a superior cap limit $\sigma(u) = 1$. For that would imply that for each constant $b > 1$, and for $a = a(u)$,

$$u \sim 0 \qquad (\text{on } F_b \text{ d-mod } F_a).$$

But $F_b = M$, and $a(u) < 1$, and it would follow from the condition of regularity that for some constant $c$ between $a$ and 1,

$$u \sim 0 \qquad (\text{on } F_c \text{ d-mod } F_a)$$

so that $\sigma(u) < 1$.

*Inferior cap limit* $\tau(u)$. A $k$-cap $u$ which is non-linkable satisfies (8.4) with each constant $b \geqq a$, as follows from Lemma 8.1. For such a $k$-cap, $\sigma(u) = \infty$. But when $u$ is non-linkable $\beta u$ is an $F$-cycle $v$ with $s(v) = a(u)$. In such a case the inferior cycle limit of $\beta u$ will be called the *inferior cap limit* of $u$, and denoted by $\tau(u)$. Thus $\tau(u) = t(v)$. The function $\tau(u)$ is not defined when $u$ is linkable.

The boundaries of non-linkable $k$-caps $u$ in the same cap class have the same superior cycle limit $a(u)$, and are in the same $F$-homology class by virtue of Lemma 8.1. They accordingly have a common inferior cap limit $\tau$. Thus inferior and superior cap limits depend only on the cap class in which a $k$-cap is found.

*The span of a k-cap.* If $u$ is linkable $k$-cap, its cap span shall be

$$S(u) = \sigma(u) - a(u).$$

If $u$ is non-linkable, its cap span shall be

$$S(u) = a(u) - \tau(u).$$

As we have seen in Example 8.1 the cap span of a $k$-cap may be null. The span of a non-linkable $k$-cap $u$ is the span of $\beta u$ as an $F$-cycle. Hence $S(u) \neq 0$.

Let $u$ and $v$ be $k$-caps. If $a(u) < a(v)$, $u + v$ is in the cap class of $v$ so that

$$S(u + v) = S(v).$$

If $a(u) = a(v)$, and $u$ is linkable but $v$ non-linkable, $u + v$ is non-linkable and

$$S(u + v) = S(v).$$

The span of an $F$-cycle $u$ with bi-lexical ranks $(s, t)$ is $s - t$. This will now be termed the *cycle span* of $u$ to distinguish it from the *cap span* $S(u)$ in case $u$ is a $k$-cap.

Let $e$ be a positive constant.

THEOREM 8.2. *The rank function $a_e(u)$ defined by the cap heights $a(u)$ of $k$-caps with cap spans $> e$ satisfies the rank conditions $R$.*

We rely on Theorem 4.1, taking $A$ as the set of all $k$-caps and $E$ as the subset of

$k$-caps with cap spans $> e$. The cap heights $a(u)$ shall replace the ranks $\rho(u)$. The set $E$ is an $a$-invariant subset of $A$ since caps in the same cap class have the same cap span. It remains to verify condition V of §4.

Condition V is concerned with $k$-caps $u$ and $v$ generated by $k$-caps $x$ of $E$ for which $a(x) \leqq \omega$. Condition V affirms that if

$$a(u) = a(v) = a(u + v) = \omega, \tag{8.5}$$

and if $S(u) \leqq e$ and $S(v) \leqq e$, then $S(u + v) \leqq e$.

To establish V we shall first show that $u$ and $v$ in V are linkable. Observe that the generating $k$-caps $x$ in V satisfy the homology

$$\beta x \sim 0 \qquad \text{(d-on } F_\omega \text{ d-mod } F_{\omega - e}\text{)}. \tag{8.6}$$

This is trivial if $x$ is linkable. If $x$ is non-linkable, it is a consequence of the fact that $S(x) > e$. It follows that $u$ and $v$ satisfy (8.6) as well as $x$. If $u$ were non-linkable, $\tau(u) < \omega - e$ since $u$ satisfies (8.6), so that

$$e < \omega - \tau(u) = S(u),$$

contrary to hypotheses in V. Hence $u$ is linkable. Similarly $v$ is linkable.

It follows from (8.5) that $u + v$ is linkable. To establish V we observe that when (8.5) holds,

$$\sigma(u + v) \leqq \max [\sigma(u), \sigma(v)]. \tag{8.7}$$

Without loss of generality we can suppose that $\sigma(u) \geqq \sigma(v)$. Then

$$S(u + v) = \sigma(u + v) - \omega \leqq \sigma(u) - \omega = S(u) \leqq e. \tag{8.8}$$

Thus V holds. Theorem 8.2 follows from Theorem 4.1.

COROLLARY 8.2. *The rank function $a_1(u)$ defined by the cap heights of $k$-caps with infinite cap spans satisfies the rank conditions R.*

This follows from the theorem upon setting $e = 1$. For the only spans which exceed 1 are infinite. In this connection recall that the only $k$-caps with infinite cap spans are linkable.

In the following theorem $a_e'(u)$ should be read with "linkable" and $a_e''(u)$ with "non-linkable."

THEOREM 8.3. *The rank function $a_e'(u)$ $[a_e''(u)]$ defined by the cap heights $a(u)$ of linkable [non-linkable] $k$-caps with finite spans $> e$ satisfies the rank conditions R.*

We again turn to Theorem 4.1, taking $A$ as the set of all $k$-caps with spans $> e$, and $E$ as the subset of linkable [non-linkable] $k$-caps with finite spans $> e$. The ranks $a_e(u)$ of the preceding theorem will serve as the ranks $\rho(u)$. The set $E$ is an $a_e$-invariant subset of $A$ since two $k$-caps in the same cap class will belong to $E$, if either belongs to $E$. It remains to verify condition V of §4.

Condition V affirms that if $u$ and $v$ are generated by $k$-caps $x$ of $E$ for which $a_e(x) \leqq \omega$, if

$$a_e(\omega) = a_e(v) = a_e(u + v) = \omega,$$

and if $u$ and $v$ are not in $E$, then $u + v$ is not in $E$. In proving V we distinguish between the two theorems which may be read from Theorem 8.3.

CASE I. Linkable $k$-caps. In this case the generating caps $x$ are linkable, and it is clear that $u$, $v$, and $u + v$ are linkable and have finite spans. Condition V will be true if whenever $u$ and $v$ are generated as in V, and $S(u) \leqq e$, $S(v) \leqq e$, then $S(u + v) \leqq e$. But this follows from (8.7) and (8.8) as in the preceding proof.

CASE II. Non-linkable $k$-caps. In this case the generating caps $x$ in V are non-linkable, and it follows from (8.6) as in the proof of Theorem 8.2 that $u$, $v$, and $u + v$ are linkable. Hence $u + v$ is not in $E$, and V is true.

Theorem 8.3 follows from Theorem 4.1.

## 9. Relation between $k$-caps and $F$-cycles

A $k$-cycle $u$ which is also a $k$-cap will be called a *$k$-cycle cap*. Each $k$-cycle cap $u$ is an $F$-cycle. Otherwise

$$u \sim 0 \qquad (\text{d-on } F_b) \tag{9.0}$$

for each positive constant $b > a(u)$. It would follow from Lemma 5.1 that $u$ would be $a$-homologous to zero, contrary to the fact that $u$ is a $k$-cap with cap height $a$. Since $u$ is both a $k$-cap and an $F$-cycle the following terms are well defined:

| | |
|---|---|
| $a(u)$, | the cap height of $u$, |
| $\sigma(u)$, | the superior cap limit of $u$, |
| $t(u)$, | the inferior cycle limit of $u$, |
| $s(u)$, | the superior cycle limit of $u$, |
| $S(u)$, | the cap span $\sigma(u) - a(u)$ of $u$, |
| $s - t$, | the cycle span of $u$. |

EXAMPLE 9.1. Let $M$ be the interval $-2\pi \leqq x \leqq 2\pi$. Let $F(x)$ be a function such that

$$7F(x) = 2\cos x + 3 \qquad (-\pi/2 \leqq x \leqq \pi/2)$$

with $7F(x) = -\cos x + 3$ on the remainder of $M$. The function $F$ is positive and continuous on $M$. Let $x$ and $y$ be Vietoris 0-cycles with the points $x = \pm \pi/2$ respectively as carriers. Let $v$ be the cycle cap $x - y$. One sees that

$$a(v) = \tfrac{3}{7}, \qquad \sigma(v) = \tfrac{4}{7}, \qquad s(v) = \tfrac{5}{7}, \qquad t(v) = \tfrac{2}{7}.$$

LEMMA 9.1. *Each $k$-cycle cap $u$ is an $F$-cycle with*

(a). $a(u) \geqq t(u), \qquad s(u) \geqq \sigma(u).$

(b). If $s(u) = \sigma(u), \qquad a(u) = t(u).$

(c). If $a(u) = t(u), \qquad s(u) = \sigma(u).$

That $u$ is an $F$-cycle has just been shown. We turn to the proof of (a).

(a). The cycle $u$ is on $F_a$ so that $a(u) \geqq t(u)$. If $\sigma(u) = a(u)$, $s(u) > \sigma(u)$ since $u$ is an $F$-cycle. We therefore suppose that $\sigma(u) > a(u)$. Let $b$ be any constant between $\sigma(u)$ and $a(u)$. It follows from the definition of $\sigma(u)$ that

$$u \nsim 0 \qquad (\text{on } F_b \text{ d-mod } F_a).$$

These constants $b$ are homology bounds of $u$ so that $s(u) \geqq \sigma(u)$ as stated in (a).

(b). If $s(u) = \sigma(u)$, $t(u) = a(u)$. For if $t(v) < a(u)$, it would follow from the definition of $t(u)$ that

$$u \sim 0 \qquad (\text{d-on } F_s \text{ d-mod } F_a).$$

But $s = \sigma$ so that this is contrary to the definition of $\sigma$. Hence $t(u) = a(u)$.

(c). If $t(u) = a(u)$, $\sigma(u) = s(u)$. For if $s(u) > \sigma(u)$,

$$u \sim 0 \qquad (\text{d-on } F_s \text{ d-mod } F_a)$$

by virtue of the definition of $\sigma(u)$. It would follow from the definition of $t(u)$ that $t(u) < a(u)$. Hence $\sigma(u) = s(u)$.

The proof of the lemma is complete.

LEMMA 9.2. *Let $\eta$ be an arbitrary positive constant. The cap class of each linkable $k$-cap $u$ contains at least one $k$-cycle cap $z$ such that*

$$s(z) \leqq \sigma(z) + \eta. \tag{9.1}$$

There are two cases according as $\sigma(u)$ is finite or infinite.

CASE I. $\sigma(u) < \infty$. Here $\sigma < 1$. By virtue of the definition of $\sigma(u)$ and of $a(u)$,

$$u \sim 0 \qquad (\text{on } F_{\sigma+\eta} \text{ d-mod } F_a).$$

Without loss of generality we can suppose that $\sigma + \eta < 1$ so that $F_{\sigma+\eta}$ is compact. It follows from Theorem 5.2 that there exists a $(k+1)$-cycle $v$ mod $F_a$ on $F_{\sigma+\eta}$ such that $\beta v - u$ is d-below $a$. Set $\beta v = z$. Then $z$ is a $k$-cycle on $F_a$. It is also a $k$-cap in the cap class of $u$ since $z - u$ is d-below $a$. Moreover $z$ is an $F$-cycle by virtue of the preceding lemma. Hence $s(z)$ exists. But $z = \beta v$ so that $z \sim 0$ on $F_{\sigma+\eta}$. Hence (9.1) holds, and the proof of the lemma is complete.

CASE II. $\sigma(u) = \infty$. Since $u$ is linkable there exists a constant $c < a$ such that

$$\beta u \sim 0 \qquad (\text{on } F_c).$$

We apply Theorem 5.3, setting $A = F_c$ and $C = F_a$. We thereby infer the existence of an absolute $k$-cycle $z$ on $F_a$ such that $u \sim z$ mod $F_c$ on $F_a$. Hence $z$ is a $k$-cycle cap in the cap class of $u$. Since $\sigma(u) = \infty$, it follows from Lemma 9.1 that $s(u) = \infty$, and (9.1) is satisfied.

THEOREM 9.1. *The cap class of each linkable $k$-cap $u$ of positive cap span contains at least one canonical $k$-cycle cap $v$. For any such cycle,*

$$t(v) = a(v), \qquad \sigma(v) = s(v). \tag{9.2}$$

CASE I. $\sigma(u) = \infty$. There exists a $k$-cycle cap $v$ in the cap class of $u$ by virtue of the preceding lemma. In this case $\sigma(v) = \infty$ so that $s(v) = \sigma(v)$. It follows from (b) in Lemma 9.1 that $t(v) = a(v)$.

CASE II. $\sigma(u) < \infty$. Let $\eta_n$ be a sequence of positive constants converging to zero. It follows from the preceding lemma that there exists a $k$-cycle cap $z_n$ in the cap class of $u$ such that

$$s(z_n) \leqq \sigma(u) + \eta_n .$$

It follows from Lemma 9.1 (a) that

$$s(z_n) - t(z_n) \geqq \sigma(u) - a(u). \tag{9.3}$$

But the left member of (9.3) is the cycle span of $z_n$ and the right member the cap span $S(u)$ of $u$. Recall that $S(u) > 0$ by hypothesis.

At most a finite number of the numbers $s(z_n)$ are different from $\sigma(u)$. Otherwise the ranks of the cycles $z_n$ would form an infinite set of distinct ranks with cycle spans at least $S(u)$. Since $t(z_n)$ does not tend to 1 as $n$ becomes infinite this is contrary to Theorem 7.6. We conclude that $s(v) = \sigma(u)$ for some cycle $v$ of the set $z_n$.

That $v$ is canonical and that (9.2) holds now follows from Lemma 9.1.

*The capping of F-cycles.* Let $u$ be an $F$-$k$-cycle with a finite superior cycle limit $s$. A $(k+1)$-cycle $v$ d-mod $F_s$ on $F_s$ such that

$$\beta v = u$$

will be said to *cap* $u$. This term is justified by the fact that when $v$ caps $u$, $v$ is a non-linkable $(k+1)$-cap with $a(v) = a(s)$. For $v$ is on $F_s$ and satisfies no $s$-homology. Otherwise it would follow from Lemma 8.1 that

$$\beta v \sim 0 \qquad \text{(d-below } s\text{)}$$

contrary to the nature of $s(u)$. Concerning the existence of a relative $(k+1)$-cycle $v$ capping $u$ we have the following theorem.

THEOREM 9.2. *Each F-k-cycle u with a finite superior limit s is capped by at least one relative* $(k+1)$*-cycle v. Moreover v is non-linkable and*

$$a(v) = s(u).$$

The limit $s$ is finite so that $s < 1$. Thus $F_s$ is compact. By virtue of the definition of $s$,

$$u \sim 0 \qquad \text{(on } F_{s+e}\text{)}$$

for each positive constant $e$. It follows from Lemma 5.1 that

$$u \sim 0 \qquad \text{(on } F_s\text{)}.$$

Moreover $u$ lies on some compact set $F_b$ for which $b < s$. We now apply Theorem 5.2, setting

$$A = 0, \qquad B = F_b , \qquad C = F_s .$$

We thereby infer the existence of a $(k+1)$-cycle $v$ mod $F_b$ on $F_a$ such that $\beta v = u$.

The first statement of the theorem follows. The concluding statement of the theorem has already been proved.

THEOREM 9.3. *The cap heights of k-caps u with cap spans* $> e > 0$ *cluster at most at the value* 1.

If $u$ is linkable, it follows from Theorem 9.1 that there exists a $k$-cycle cap $v$ with rank $[\sigma(u), a(u)]$ and accordingly with cycle span

$$\sigma(u) - a(u) > e.$$

It follows from Theorem 7.6 that the cap heights $a(u)$ of this type cluster at most at the value 1.

If $u$ is non-linkable, $\beta u$ is a $(k-1)$-cycle $v$ with rank $[a(u), \tau(u)]$ and accordingly with cycle span

$$\tau(u) - a(u) > e \qquad [\tau(u) < a(u) < 1].$$

It follows again from Theorem 7.6 that the values $\tau(u)$ and hence the values $a(u)$ cluster at most at the value 1.

THEOREM 9.4. *The cap heights of k-caps u with finite cap spans* $> e > 0$ *are finite in number.*

The proof of this theorem is similar to the proof of Theorem 9.3 making use of Lemma 7.4.

## 10. Cap-isomorphisms and $F$-isomorphisms

Recall that a canonical $k$-cycle $u$ is an $F$-cycle cap such that $t(u) = a(u)$. We continue with the following lemma.

LEMMA 10.1 *The cap span of a canonical k-cycle u equals its cycle span. The cap span of the sum,*

$$v = v_1 + \cdots + v_n,$$

*of a finite set of canonical k-cycles* $v_i$ *with increasing ranks* $t(v_i)$ *is at least the cycle span of* $v_n$.

If $u$ is canonical $t(u) = a(u)$, and then $s(u) = \sigma(u)$ by virtue of Lemma 9.1 (a). Thus

$$S(u) = \sigma(u) - a(u) = s(u) - t(u),$$

and the first statement of the lemma is true.

To prove the second statement of the lemma set $t_i = t(v_i)$, $i = 1, \cdots, n$, and recall that $v_i$ is on $F_{t_i}$ since $v_i$ is canonical. Observe that $v$ is a $k$-cap with $a(v) = t_n$. Let $b$ be any constant between $t_n$ and $\sigma(v_n)$. For $i = 1, 2, \cdots, n-1$,

$$v_i \sim 0 \qquad (\text{on } F_b \text{ d-mod } F_{t_n})$$

since $v_i$ is on $F_{t_i}$ and $t_i < t_n$. Moreover

$$v_n \nsim 0 \qquad (\text{on } F_b \text{ d-mod } F_{t_n}) \tag{10.1}$$

by virtue of the choice of $b < \sigma(v_n)$ and the definition of $\sigma(v_n)$. Hence (10.1) holds for $v$ as well as for $v_n$. Taking account of the arbitrariness of $b$ between $t_n$ and $\sigma(v_n)$ we see that $\sigma(v) \geqq \sigma(v_n)$. Hence

$$\sigma(v) - a(v) \geqq \sigma(v_n) - a(v_n) = s(v_n) - t(v_n).$$

Thus $S(v)$ is at least the cycle span of $v_n$, and the proof of the lemma is complete.

Recall that two $F$-cycles are in the same $F$-homology class if in the same $s$-class. Here $s(u)$ is the rank function defined by the superior cycle limit of $u$. An $s$-isomorphism will be termed an *$F$-isomorphism*. Recall that two $F$-cycles which are in the same $F$-homology class are in the same $r$-class, but not conversely. Here $r(u)$ is the rank function defined by $[s(u), t(u)]$.

THEOREM 10.1. *Each group $g$ of (bounding) $k$-cycles with cycle spans $> e > 0$ if $F$-isomorphic with a group $\gamma$ of $k$-cycle caps with (finite) cap spans $> e > 0$.*

We shall make use of Theorem 2.7. To that end let $(u)$ be an $r$-proper base for $g$. That $g$ possesses an $r$-proper base follows from the fact (see Theorem 7.6) that its ranks $r$ form a discrete ascending sequence. But an $r$-proper base is at once seen to be an $s$-proper base since the ranks $r = (s, t)$ are ordered lexicographically. To apply Theorem 2.7 we map each element $u_i$ of the base $(u)$ onto a canonical $k$-cycle $v_i$ in the $F$-homology class of $u_i$. This is possible by virtue of Theorem 7.2. This mapping will be isomorphically extended to the group $\gamma$ defined by the base $(v)$. So extended the mapping defines an $F$-isomorphism of $g$ with $\gamma$.

It remains to show that the non-null cycles $z$ of $\gamma$ are $k$-caps with (finite) cap spans $> e$. Let $x_1, \cdots, x_m$ be distinct elements in the base $(v)$, and let $\delta_1, \cdots, \delta_m$ be arbitrary non-null elements of $\Delta$. We distinguish between two cases.

CASE I. $z = \delta_i x_i$ with $t(x_1) = \cdots = t(x_m) = \text{const.} = a$.

CASE II. Not Case I.

In Case I let

$$s = \max [s(x_1), \cdots, s(x_m)].$$

Then $r(z) = (s, a)$ since $(u)$ is an $r$-proper base. In particular $t(z) = a$. But $z$ is on $F_a$. Thus $z$ is canonical so that its cycle span $s - a$ equals its cap span $S(z)$. In the special case in which the $F$-cycles of $g$ are bounding, $s - a$ is finite. Hence $S(z)$ is finite in this special case.

In Case II, $z$ has the form

$$z = y_1 + \cdots + y_n, \qquad (n > 1)$$

where $y_i$ is a canonical $k$-cycle of $\gamma$ of the type of $z$ in Case I and the ranks $t_i = t(v_i)$ are different. Suppose that $t_i$ increases with $i$. It follows that $z$ is a $k$-cap with cap height $t_n$, and we infer from Lemma 10.1 that the cap span of $z$ is at least the cycle span of $v_n$, that is exceeds $e$.

In the special case in which the $F$-cycles of $g$ are bounding, we must show that

We thereby infer the existence of a $(k+1)$-cycle $v$ mod $F_b$ on $F_s$ such that $\beta v = u$.

The first statement of the theorem follows. The concluding statement of the theorem has already been proved.

THEOREM 9.3. *The cap heights of k-caps u with cap spans* $> e > 0$ *cluster at most at the value* 1.

If $u$ is linkable, it follows from Theorem 9.1 that there exists a $k$-cycle cap $v$ with rank $[\sigma(u), a(u)]$ and accordingly with cycle span

$$\sigma(u) - a(u) > e.$$

It follows from Theorem 7.6 that the cap heights $a(u)$ of this type cluster at most at the value 1.

If $u$ is non-linkable, $\beta u$ is a $(k-1)$-cycle $v$ with rank $[a(u), \tau(u)]$ and accordingly with cycle span

$$\tau(u) - a(u) > e \qquad [\tau(u) < a(u) < 1].$$

It follows again from Theorem 7.6 that the values $\tau(u)$ and hence the values $a(u)$ cluster at most at the value 1.

THEOREM 9.4. *The cap heights of k-caps u with finite cap spans* $> e > 0$ *are finite in number.*

The proof of this theorem is similar to the proof of Theorem 9.3 making use of Lemma 7.4.

## 10. Cap-isomorphisms and $F$-isomorphisms

Recall that a canonical $k$-cycle $u$ is an $F$-cycle cap such that $t(u) = a(u)$. We continue with the following lemma.

LEMMA 10.1 *The cap span of a canonical k-cycle u equals its cycle span. The cap span of the sum,*

$$v = v_1 + \cdots + v_n,$$

*of a finite set of canonical k-cycles* $v_i$ *with increasing ranks* $t(v_i)$ *is at least the cycle span of* $v_n$.

If $u$ is canonical $t(u) = a(u)$, and then $s(u) = \sigma(u)$ by virtue of Lemma 9.1 (a). Thus

$$S(u) = \sigma(u) - a(u) = s(u) - t(u),$$

and the first statement of the lemma is true.

To prove the second statement of the lemma set $t_i = t(v_i)$, $i = 1, \cdots, n$, and recall that $v_i$ is on $F_{t_i}$ since $v_i$ is canonical. Observe that $v$ is a $k$-cap with $a(v) = t_n$. Let $b$ be any constant between $t_n$ and $\sigma(v_n)$. For $i = 1, 2, \cdots, n-1$,

$$v_i \sim 0 \qquad (\text{on } F_b \text{ d-mod } F_{t_n})$$

since $v_i$ is on $F_{t_i}$ and $t_i < t_n$. Moreover

$$v_n \nsim 0 \qquad (\text{on } F_b \text{ d-mod } F_{t_n}) \tag{10.1}$$

by virtue of the choice of $b < \sigma(v_n)$ and the definition of $\sigma(v_n)$. Hence (10.1) holds for $v$ as well as for $v_n$. Taking account of the arbitrariness of $b$ between $t_n$ and $\sigma(v_n)$ we see that $\sigma(v) \geqq \sigma(v_n)$. Hence

$$\sigma(v) - a(v) \geqq \sigma(v_n) - a(v_n) = s(v_n) - t(v_n).$$

Thus $S(v)$ is at least the cycle span of $v_n$, and the proof of the lemma is complete.

Recall that two $F$-cycles are in the same $F$-homology class if in the same $s$-class. Here $s(u)$ is the rank function defined by the superior cycle limit of $u$. An $s$-isomorphism will be termed an *$F$-isomorphism*. Recall that two $F$-cycles which are in the same $F$-homology class are in the same $r$-class, but not conversely. Here $r(u)$ is the rank function defined by $[s(u), t(u)]$.

THEOREM 10.1. *Each group $g$ of (bounding) $k$-cycles with cycle spans $> e > 0$ if $F$-isomorphic with a group $\gamma$ of $k$-cycle caps with (finite) cap spans $> e > 0$.*

We shall make use of Theorem 2.7. To that end let $(u)$ be an $r$-proper base for $g$. That $g$ possesses an $r$-proper base follows from the fact (see Theorem 7.6) that its ranks $r$ form a discrete ascending sequence. But an $r$-proper base is at once seen to be an $s$-proper base since the ranks $r = (s, t)$ are ordered lexicographically. To apply Theorem 2.7 we map each element $u_i$ of the base $(u)$ onto a canonical $k$-cycle $v_i$ in the $F$-homology class of $u_i$. This is possible by virtue of Theorem 7.2. This mapping will be isomorphically extended to the group $\gamma$ defined by the base $(v)$. So extended the mapping defines an $F$-isomorphism of $g$ with $\gamma$.

It remains to show that the non-null cycles $z$ of $\gamma$ are $k$-caps with (finite) cap spans $> e$. Let $x_1, \cdots, x_m$ be distinct elements in the base $(v)$, and let $\delta_1, \cdots, \delta_m$ be arbitrary non-null elements of $\Delta$. We distinguish between two cases.

CASE I. $z = \delta_i x_i$ with $t(x_1) = \cdots = t(x_m) = \text{const.} = a$.

CASE II. Not Case I.

In Case I let

$$s = \max [s(x_1), \cdots, s(x_m)].$$

Then $r(z) = (s, a)$ since $(u)$ is an $r$-proper base. In particular $t(z) = a$. But $z$ is on $F_a$. Thus $z$ is canonical so that its cycle span $s - a$ equals its cap span $S(z)$. In the special case in which the $F$-cycles of $g$ are bounding, $s - a$ is finite. Hence $S(z)$ is finite in this special case.

In Case II, $z$ has the form

$$z = y_1 + \cdots + y_n, \qquad (n > 1)$$

where $y_i$ is a canonical $k$-cycle of $\gamma$ of the type of $z$ in Case I and the ranks $t_i = t(v_i)$ are different. Suppose that $t_i$ increases with $i$. It follows that $z$ is a $k$-cap with cap height $t_n$, and we infer from Lemma 10.1 that the cap span of $z$ is at least the cycle span of $v_n$, that is exceeds $e$.

In the special case in which the $F$-cycles of $g$ are bounding, we must show that

$S(z)$ is finite. But in this case the cycle span of $y_i$ is finite. Thus $y_i$ is bounding for each $i$. Hence $z$ is bounding so that $s(z) - t(z)$ is finite. But

$$S(z) = \sigma(z) - a(z) \leqq s(z) - t(z).$$

We conclude that $S(z)$ is finite.

The proof of the theorem is complete.

We refer to the ranks $a(u)$ defined by the cap heights of $k$-caps $u$. An $a$-isomorphism will be called a *cap-isomorphism.* Under an $a$-isomorphism elements correspond which are in the same cap class.

THEOREM 10.2. *Each group $g$ of linkable $k$-caps $u$ with (finite) cap spans $> e > 0$ is cap-isomorphic with a group of (bounding) $F$-cycle caps with cycle spans $> e > 0$.*

The different cap heights of $k$-caps with spans $> e > 0$ form a discrete ascending sequence by virtue of Theorem 9.3. It follows from Lemma 2.3 that $g$ admits an $a$-proper base $(u)$. We map each $k$-cap $u_i$ of $(u)$ on a $k$-cycle cap $v_i$ in the same cap class with a cycle span equal to $S(u_i)$. This is possible by virtue of Theorem 9.1. By virtue of Theorem 2.7 this mapping can be extended isomorphically so as to give a cap-isomorphism of $g$ with the group $\gamma$ generated by the set $(v)$.

The cycle span of a $k$-cycle cap is at least its cap span in accordance with Lemma 9.1 (a), so that the cycle spans of $\gamma$ exceed $e$. Moreover if the cap spans of $g$ are finite, $S(u_i)$ in particular is finite. But the cycle span of $v_i$ equals $S(u_i)$ so that $v_i$ is bounding. Hence $\gamma$ is a group of bounding $k$-cycles if the cap spans of $g$ are finite.

The proof of the theorem is complete.

COROLLARY 10.2. *The dimension of a maximal group of linkable $k$-caps with finite cap spans $> e > 0$ is finite.*

This follows from Theorem 10.2 and Theorem 7.5 which states that the dimension of a maximal group of bounding $F$-cycles with spans $> e > 0$ is finite.

*$\beta$-isomorphisms.* Let $g$ be a group of formal $k$-chains $u$. The group of boundary chains $\beta u$ will be denoted by $\beta g$. The groups $g$ and $\beta g$ will be said to be $\beta$-isomorphic if $g$ and $\beta g$ are isomorphic with $u$ in $g$ corresponding to $\beta u$ in $\beta g$.

THEOREM 10.3. *A group $g$ of non-linkable $k$-caps $u$ is $\beta$-isomorphic with its boundary group $\beta g$. The cap span of a $k$-cap $u$ in $g$ equals the cycle span of its image $\beta u$ in $\beta g$.*

If $u$ in $g$ corresponds to $\beta u$ in $\beta g$, $g$ is mapped homeomorphically on $\beta g$. This homeomorphism is an isomorphism if whenever $\beta u = 0$, $u = 0$. If $\beta u = 0$ with $u \neq 0$, $u$ would be linkable contrary to the nature of $g$. Hence $u = 0$ when $\beta u = 0$, and $g$ is isomorphic with $\beta g$.

If $u \neq 0$,

$$s(\beta u) = a(u), \qquad t(\beta u) = \tau(u),$$

so that the cap span $a(u) - \tau(u)$ equals the cycle span

$$s(\beta u) - t(\beta u)$$

of $\beta u$. This completes the proof of the theorem.

COROLLARY 10.3. *The dimension of a maximal group $g$ of non-linkable $k$-caps with spans $> e > 0$ is finite.*

For the isomorphic group $\beta g$ is a group of bounding $F$-cycles with cycle spans $> e$, and has a finite dimension in accordance with Theorem 7.5.

THEOREM 10.4. *Each group $\gamma$ of bounding $F$-$k$-cycles with cycle spans $> e > 0$ is the boundary group $\beta g$ of a group $g$ of non-linkable $k$-caps with spans $> e$.*

The elements $u$ of $\gamma$ have ranks $s(u)$. The dimension of $\gamma$ is finite by virtue of Theorem 7.5 so that the ranks $s$ of $\gamma$ are finite in number. It follows from Lemma 2.3 that $\gamma$ has an $s$-proper base $(x)$. Let $x$ be an element in $(x)$. The $F$-cycle $x$ is capped by a relative $(k+1)$-cycle $v$ on $F \leqq s(x)$ so that $\beta v = x$ in accordance with Theorem 9.2. Let $(v)$ be the set of relative $(k+1)$-cycles thereby capping the cycles $x$ of $(x)$. Let $g$ be the group generated by the elements of $(v)$. It is clear that $\beta g = \gamma$.

Let $(x_1, \cdots, x_r)$ be distinct cycles in $(x)$ capped respectively by relative $(k+1)$-cycles $(v_1, \cdots, v_r)$ of $(v)$. Set

$$u = \delta_k x_k, \qquad v = \delta_k v_k \qquad (\delta_k \neq 0; k = 1, \cdots, r).$$

We understand that $u$ in $\gamma$ is mapped on $v$ in $g$. I say that $v$ is a non-linkable $k$-cap capping $u$. For

$$s(u) = \max [s(x_1), \cdots, s(x_r)]$$

since $(x)$ is an $s$-proper base. Moreover $v$ is on $F \leqq s(u)$ and $\beta v = u$. Thus $v$ caps $u$ by definition.

The cap span of $v$ is

$$a(v) - \tau(v) = s(u) - t(v).$$

But $s(u) - t(u) > e$ by hypothesis, and the proof of the theorem is complete.

## 11. Maximal groups

We shall be concerned with the following groups ($k \geqq 0$, $e > 0$):

$F_k^e$, a maximal group of bounding $F$-cycles with cycle spans $> e$,

$\Lambda_k^e$, a maximal group of linkable $k$-caps with finite cap spans $> e$,

$N_k^e$, a maximal group of non-linkable $k$-caps with cap spans $> e$,

$C_k^e$, a maximal group of $k$-caps with finite cap spans $> e$.

The dimensions of the first three of these groups are finite in accordance with Theorem 7.5, and Corollaries 10.2 and 10.3. And as we have seen, rank functions $r_e(u)$, $a_e'(u)$, $a_e''(u)$ can be respectively defined for elements $u$ of the type appearing in these three groups, and so defined satisfy the rank conditions $R$. Theorem 2.8 accordingly yields the following.

THEOREM 11.1. *Any group $F_k^e$, $\Lambda_k^e$, or $N_k^e$ is $\rho$-isomorphic with any other*

*group $F_k^e$, $\Lambda_k^e$, or $N_k^e$ respectively, provided $\rho(u)$ represents the rank function appropriate to elements in the respective groups.*

COROLLARY 11.1. *The dimensions of maximal groups of the above types are independent of the choice of the maximal groups among groups of the same types.*

LEMMA 11.1. *The groups $F_k^e$ and $\Lambda_k^e$ have the same dimension.*

By virtue of Theorem 10.1, $F_k^e$ is $F$-isomorphic with a group of $k$-caps of the type appearing in $\Lambda_k^e$ so that

$$\dim F_k^e \leqq \dim \Lambda_k^e .$$

On the other hand it follows from Theorem 10.2 that $\Lambda_k^e$ is cap-isomorphic with a group of $F$-cycles of the type appearing in $F_k^e$ so that

$$\dim F_k^e \geqq \dim \Lambda_k^e .$$

The lemma follows directly.

LEMMA 11.2. *The groups $N_{k+1}^e$ and $F_k^e$ have the same dimension.*

The proof of this lemma is similar to that of the preceding. It depends upon Theorems 10.3 and 10.4.

THEOREM 11.2. *There exists a maximal group $N_{k+1}^e$ of non-linkable $(k+1)$-caps with spans $> e > 0$, such that $\beta N_{k+1}^e$ is a group $F_k^e$ and a group $\Lambda_k^e$.*

Starting with an arbitrary group $F_k^e$, Theorem 10.1 affirms the existence of an $F$-isomorphic group $\gamma_k^e$ of bounding $k$-cycle caps with cap spans $> e$. The group $\gamma_k^e$ will be a group $\Lambda_k^e$ if it has the dimension of $\Lambda_k^e$. That this is the case now follows from Lemma 11.1. The group $\gamma_k^e$ is thus a group $F_k^e$ and a group $\Lambda_k^e$.

It follows from Theorem 10.4 that $\gamma_k^e$ is the boundary group $\beta M_{k+1}^e$ of a group of non-linkable $(k+1)$-caps with spans $> e$. The group $M_{k+1}^e$ will be a group $N_{k+1}^e$ if it has the dimension of $N_{k+1}^e$. That this is the case follows from Lemma 11.2. Hence $M_{k+1}^e$ is a group $N_{k+1}^e$. For such a choice of $N_{k+1}^e$, the group $\beta N_{k+1}^e$ is a group $F_k^e$ and a group $\Lambda_k^e$ as stated.

The proof of the theorem is complete.

THEOREM 11.3. *Each group $C_k^e$ is a direct sum of the form*

$$C_k^e = \Lambda_k^e + N_k^e , \tag{11.1}$$

*where $\Lambda_k^e$ and $N_k^e$ are suitably chosen groups of their respective types.*

Let $a^*(u)$ be the rank function defined by the cap heights of elements of $C_k^e$. The function $a^*(u)$ satisfies conditions $R$ by virtue of Theorem 2.9. We shall now apply Theorem 4.2, taking $a^*(u)$ as the rank function $\rho$. Let $A$ be the set of $k$-caps in $C_k^e$. Let $B$ and $C$ be respectively the $a^*$-invariant subsets of linkable and non-linkable $k$-caps of $A$.

Conditions (i), (ii), and (iii) preceding Theorem 4.2 here take the form:

(i). The sum of two linkable $k$-caps in $A$ in different cap classes is a linkable $k$-cap in $A$.

(ii). The sum of two non-linkable $k$-caps in $A$ with different cap heights is a non-linkable $k$-cap in $A$.

(iii). The sum of a linkable $k$-cap and a non-linkable $k$-cap in $A$ is in $A$.

These conditions are obviously fulfilled. It follows from Theorem 4.2 that $C_k^e$ has the form

$$C_k^e = b + c,$$

where $b$ and $c$ are respectively maximal groups of elements in $B$ and $C$. But $b$ and $c$ must be groups $\Lambda_k^e$ and $N_k^e$ respectively. Otherwise one of the groups $b$ and $c$, say $b$, will be a proper subgroup of a group $b'$ of the type $\Lambda_k^e$. The group $b' + c$ would then be of the type of $C_k^e$, and contain a group of the type $C_k^e$ as a proper subgroup. This is impossible since each group $C_k^e$ is maximal. Thus $b$ is a group $\Lambda_k^e$. Similarly $c$ is a group $N_k^e$.

The proof of the theorem is complete.

COROLLARY 11.3. *For a fixed $k$ and $e > 0$, the dimensions of all groups of the type $C_k^e$ are the same.*

This is true of groups $C_k^e$ because each group $C_k^e$ has the form (11.1) and groups of the types $\Lambda_k^e$ and $N_k^e$ have dimensions independent of their mode of formation.

THEOREM 11.4. *If $N_k^e$ is the maximal group of non-linkable $k$-caps of span $> e$ affirmed to exist in Theorem* 11.2, *the group*

$$N_k^e + \beta N_{k+1}^e \qquad (k = 0, 1, \cdots) \tag{11.2}$$

*is a maximal group $C_k^e$ of $k$-caps with finite spans $> e$.*

By virtue of the choice of the group $N_{k+1}^e$, the group (11.2) is of the form

$$N_k^e + \Lambda_k^e.$$

As such it is clearly a group of $k$-caps with finite spans $> e$. It will be a group $C_k^e$ if its dimension is that of a group $C_k^e$. That this is the case follows from Theorem 11.3 and its corollary.

Consider the special case where $0 = R_1 = R_2 = \cdots$. For $k > 0$ there will then be no non-bounding $k$-cycles. There can then be no linkable $k$-cap $u$ with an infinite cap span. For there is an $F$-cycle cap $v$ in the cap class of $u$, and $s(v) \geqq \sigma(u)$. If $\sigma(u) = \infty$, $s(v) = \infty$ and $v$ is non-bounding, contrary to hypothesis. If then $0 = R_1 = R_2 = \cdots$, the word finite can be deleted from the preceding theorem for $k > 0$, giving the following corollary.

COROLLARY 11.4. *In the case in which $R_1 = R_2 = \cdots = 0$, the groups*

$$C_k^e = N_k^e + \beta N_{k+1}^e \qquad (k = 1, 2, \cdots) \tag{11.3}$$

*of the theorem are maximal groups of $k$-caps with spans $> e$.*

The case $0 = R_1 = R_2 = \cdots$ is particularly important. It occurs in the theory of minimal surfaces spanning a simple closed curve.

Let $c_k^e$ and $n_k^e$ be respectively the dimensions of the groups $C_k^e$ and $N_k^e$ of the theorem. It follows from the theorem that except for $k = 0$,

$$c_k^e = n_k^e + n_{k+1}^e \qquad (k = 0, 1, \cdots). \tag{11.4}$$

The dimension $n_0^e = 0$ since there are no non-linkable 0-caps. A necessary and sufficient condition that a sequence of non-negative integers $c_0^e$, $c_1^e$, $\cdots$ be representable in the form (11.4) with each $n_k^e \geqq 0$ is that

$$c_m^e - c_{m-1}^e + \cdots + (-1)^m c_0^e \geqq 0 \qquad (m = 0, 1, \cdots). \tag{11.5}$$

For when the numbers $c_k^e$ are representable in the form (11.4) we find that the left member of (11.5) equals $n_{m+1}^e$ so that (11.5) holds. Conversely if (11.5) holds for each $m$ and we define $n_{m+1}^e$ as the left member of (11.5) with $n_0^e = 0$, we find that (11.4) holds. Thus (11.4) and (11.5) may be regarded as equivalent sets of conditions on the numbers $c_k^e$.

Many other interesting relations follow from (11.4), in particular the relations

$$c_{m+1}^e + c_{m-1}^e \geqq c_m^e \qquad (m = 1, 2, \cdots). \tag{11.6}$$

In the theory of critical points of functions, (11.6) would lead to the following theorem. Among critical points "of span" $e > 0$, the "number" of critical points of types $m + 1$ and $m - 1$ is at least the number of critical points of type $m$. This is for the special case in which $0 = R_1 = R_2 = \cdots$. We now return to the more general case.

## 12. Type number relations

We were concerned in the preceding section with $k$-caps with finite cap spans. We now bring in $k$-caps of infinite cap span. Such $k$-caps $u$ are linkable with $\sigma(u) = \infty$.

THEOREM 12.1. *Each maximal group $P_k$ of $k$-caps with infinite cap spans is cap-isomorphic with a group $\gamma$ of canonical non-bounding $k$-cycles containing one and only one $k$-cycle in each homology class.*

Any cap span which is at least 1 is infinite. Hence $P_k$ is a maximal group of $k$-caps with cap spans $> 1$. It follows from Theorem 10.2 that $P_k$ is cap-isomorphic with a group of $k$-cycle caps with cycle spans $> 1$. The cycles $v$ of $\gamma$ accordingly have infinite cycle spans and so are non-bounding. Thus $s(v) = \sigma(v) = \infty$. It follows from Lemma 9.1 that $t(v) = a(v)$ so that $v$ is canonical.

Since the cycles of $\gamma$ are non-bounding $\gamma$ contains at most one cycle in each homology class. Let $H$ be the set of homology classes which contain no cycle of $\gamma$. It remains to show that the set $H$ is empty.

Let $r_0$ be the minimum rank of cycles in classes of $H$. Such a minimum exists by virtue of Theorem 7.6. Let $z$ be a canonical $k$-cycle with rank $r_0$ and in some class of $H$.

Since $P_k$ is a maximal group of $k$-caps with infinite cap spans there exists a $k$-cap $y$ in $P_k$ such that $z - y$ is not a $k$-cap with infinite cap span. Setting $a = a(z)$ we infer that

$$z - y \sim 0 \qquad (\text{d-mod } F_a).$$

Let $x$ be the image of $y$ in $\gamma$. Since $x$ and $y$ are $a$-homologous

$$z - x \sim 0 \qquad \text{(d-mod } F_a\text{)}. \tag{12.1}$$

The cycle $x$ is in an homology class of $\gamma$. The cycle $z - x$ is in no such class since $z$ is in no such class. Hence $z - x$ is non-bounding. Thus $s(z - x) = \infty$, and it follows from (12.1) that $t(z - x) < a$. Hence

$$r(z - x) < r(z).$$

This is contrary to the choice of $r_0 = r(z)$ as the minimum of the ranks of cycles in classes of $H$. We infer that $H$ is empty.

The proof of the theorem is complete.

COROLLARY 12.1. *If the $k^{\text{th}}$ connectivity $p_k$ is infinite, there exist infinitely many distinct cap heights belonging to $k$-caps with infinite cap spans. These cap heights cluster only at the value $a = 1$.*

Let $(\infty, t_n)$ be the ranks $r$ of a base of the group $\gamma$ of the theorem. If the cap heights $t_n$ assume at most a finite set of values let $b$ be the maximum of such values. Then the cycles of $\gamma$ lie on $F_b$. If $c$ is a constant such that $b < c < 1$ the connectivities $R(b, c)$ would then be infinite contrary to fact.

We infer that $t_n$ assumes infinitely many different values. It follows from Theorem 7.6 that Lim $t_n = 1$, and the proof of the corollary is complete.

THEOREM 12.2. *Let $C_k^e$ and $P_k$ be respectively maximal groups of $k$-caps with finite cap spans $> e > 0$ and with infinite cap spans. The sum*

$$g = C_k^e + P_k \tag{12.2}$$

*is a maximal group of $k$-caps with spans $> e$.*

The group $g$ is composed of $k$-caps with spans $> e$. Let $g^*$ be a maximal group of $k$-caps with spans $> e$, containing $g$. We shall prove that $g^* = g$.

To that end we apply Theorem 4.2, taking our ranks as the cap heights of caps of $g^*$, and identifying the set $A$ with $g^*$. The sets $B$ and $C$ shall consist respectively of $k$-caps of $g^*$ with finite and with infinite spans. Conditions (i), (ii), and (iii) of Theorem 4.2 are immediately verified. Moreover the different ranks of elements of $C$ form an ascending sequence. It follows from Theorem 4.2 that

$$g^* = C_k^e + P_k,$$

and we conclude that $g = g^*$.

The proof of the theorem is complete.

Theorems 11.4 and 12.2 lead to the following corollary.

COROLLARY 12.2. *If $N_k^e$ is a suitably chosen maximal group of non-linkable $k$-caps with cap spans $> e > 0$, then*

$$M_k^e = N_k^e + \beta N_{k+1}^e + P_k \qquad (k = 0, 1, \cdots)$$

*is a maximal group of $k$-caps with cap spans $> e$.*

Corresponding to each cap height $a$, the dimension of a maximal group of

$k$-caps with cap heights $a$ and with spans $> e$ will be called the $k^{\text{th}}$ *e-type number* $m_k^e(a)$ of $a$. The rank function $a_e(u)$ defined by cap heights of $k$-caps with cap spans $> e$ satisfies conditions $R$. It follows from Theorem 2.1 that the number $m_k^e(a)$ is independent of the particular maximal group chosen. We set

$$m_k^e = \sum_a m_k^e(a),$$

where $a$ ranges over all cap heights and term $m_k^e$ the $k^{\text{th}}$ *e-type number sum.*

The cap heights $a_e(u)$ form a discrete ascending sequence and satisfy the rank conditions $R$. Hence $M_k^e$ is the direct sum of maximal groups of $k$-caps with the respective cap heights $a_e(u)$ and with cap spans $> e$. It follows that

$$\dim M_k^e = m_k^e .$$

If we set

$$\dim N_k^e = n_k^e , \qquad \dim P_k = p_k ,$$

we have the following theorem.

THEOREM 12.3. *For $e > 0$, the $k^{\text{th}}$ e-type number sums $m_k^e$ are given by relations of the form*

$$m_k^e = n_k^e + n_{k+1}^e + p_k \qquad (k = 0, 1, \cdots). \tag{12.3}$$

The numbers $n_k^e$ are finite and $n_0^e = 0$. The numbers $p_k$ are the connectivities of $M$ in accordance with Theorem 12.1 and may be infinite. If $p_k$ is infinite, the right member of (12.3) is regarded as infinite.

COROLLARY 12.3. *If the connectivities $p_k$ are finite and if we set $\mu_k = m_k^e - p_k$, then*

$$\mu_n - \mu_{n-1} + \cdots + (-1)^n \mu_0 \geqq 0 \qquad (n = 0, 1, \cdots).$$

COROLLARY 12.4. *If for a particular $m$, the connectivities $p_{m-1}$, $p_m$, and $p_{m+1}$ are finite,*

$$\mu_{m+1} + \mu_{m-1} \geqq \mu_m \qquad (m > 0).$$

COROLLARY 12.5. *The numbers $m_k^1$ give the type number sums corresponding to k-caps with infinite cap spans and are such that*

$$m_k^1 \geqq p_k \qquad (k = 0, 1, \cdots).$$

COROLLARY 12.6. *If the connectivities $p_0, \cdots, p_{m-1}$ are finite and $\mu_{m+1} = 0$,*

$$\begin{aligned} \mu_0 &\geqq 0, \\ \mu_1 - \mu_0 &\geqq 0, \\ &\cdots\cdots\cdots \\ \mu_m - \mu_{m-1} + \cdots + (-1)^m \mu_0 &= 0. \end{aligned}$$

The conditions of this corollary are satisfied if the underlying space is a finite $m$-dimensional complex as is readily seen.

We call attention to the following important corollaries.

COROLLARY 12.7. *If for some positive $e$ every cap of span $> e$ is linkable, then $m_k^e$ is the $k^{\text{th}}$ connectivity of $M$.*

COROLLARY 12.8. *If every cap is linkable, then $m_k^e$ is independent of $e$ and equals the $k^{\text{th}}$ connectivity of $M$.*

If every cap is linkable $m_k^e = m_k^\infty$, so that every cap has an infinite cap span. The connectivities then take the form

$$p_k = \sum_c m_k(c),$$

where $m_k(c)$ is the dimension of a maximal group of linkable $k$-caps with cap heights $c$. This mode of determination of $p_k$ has sufficed in certain applications where all other methods failed.

## 13. $k$-Cap heights finite in number

The dimension $m_k(c)$ of a maximal group of $k$-caps with cap heights $c$ will be called the $k^{\text{th}}$ *type number* of $c$.

THEOREM 13.1. *If a $k$-cap height $c$ is bounded from $(k-1)$- and $(k+1)$-cap heights other than $c$, the $k^{\text{th}}$ type number of $c$ is finite, and no $k$-cap $u$ with cap height $c$ has a cap span 0.*

Let $\lambda_k$ be a group of linkable $k$-caps with cap heights $c$. If dim $\lambda_k$ is finite, one sees that $\lambda_k$ is cap-isomorphic with a group of $k$-cycle caps $u$. In accordance with Lemma 9.1 each such $k$-cycle $u$ is an $F$-cycle with

$$t(u) \leqq c, \qquad s(u) \geqq c. \tag{13.0}$$

Moreover $s(u) > c$. Otherwise $t(u) = c = s(u)$ in accordance with Lemma 9.1 (b) and $u$ would not be an $F$-cycle.

Suppose there are no $(k+1)$- or $(k-1)$-cap heights on the closed interval $(c-e, c+e)$, possibly excepting $c$. Then $s(u) > c+e$, since $s(u)$ is a $(k+1)$-cap height. The cycles $u$ are thus on $F_c$ and non-bounding on $F_{c+e}$. Hence

$$\dim \lambda_k \leqq R\,(c, c+e). \tag{13.1}$$

Let $\omega_k$ be a group of non-linkable $k$-caps with cap heights $c$. Then $\beta\omega_k$ is an isomorphic group of $(k-1)$-cycles $v$ with $s(v) = c$. Moreover $t(v) < c - e$ since the interval $c - e \leqq t < c$ contains no $(k-1)$-cap heights. Let $b$ be a constant between $c-e$ and $c$. Then $v$ is on $F_{c-e}$ and non-bounding on $F_b$. Hence

$$\dim \omega_k \leqq R(c-e, b). \tag{13.2}$$

If $\lambda_k$ and $\omega_k$ are maximal, one sees that $\lambda_k + \omega_k$ is a maximal group of $k$-caps with cap heights $c$. But the dimension of $\lambda_k + \omega_k$ will be finite in accordance with (13.1) and (13.2). The first affirmation of the theorem is accordingly proved.

No $k$-cap $u$ with $a(u) = c$ has a cap span 0. For if $S(u) = 0$, $u$ would be

linkable. It would follow from Lemma 9.2 that the cap class of $u$ would contain at least one $k$-cycle cap $z$ such that

$$s(z) < c + e,$$

with $c < s(z)$ as we have seen in connection with (13.0). But this is contrary to the nature of $e$ since there are no $(k+1)$-cap heights on the open interval $(c, c+e)$. Hence no $k$-cap $u$ with cap height $c$ has a cap span 0.

The proof of the theorem is complete.

LEMMA 13.1. *If the $(k-1)$-, $(k)$- and $(k+1)$-cap heights are finite in number, the cap spans of $k$-caps are bounded from zero.*

There will be no $k$-caps with cap spans 0 by virtue of the preceding theorem. Let $e$ be a positive constant less than the absolute difference between any $(j+1)$-cap height and distinct $j$-cap height, for $j = k$ and $k-1$. Then the cap span of each $k$-cap is readily seen to exceed $e$.

The sum of the $k^{\text{th}}$ type numbers $m_k(c)$ of the various $k$-cap heights $c$ will be called the $k^{\text{th}}$ *type number sum.*

THEOREM 13.2. *If the $k$-cap heights are finite in number for each $k$, the $k^{\text{th}}$ type number sums $m_k$ and the $k^{\text{th}}$ connectivities $p_k$ are finite. Moreover the differences $\mu_k = m_k - p_k$ satisfy the relations*

$$\mu_n - \mu_{n-1} + \cdots + (-1)^n \mu_0 \geqq 0 \qquad (n = 0, 1, \cdots). \tag{13.3}$$

For a given $k$ and $c$, $m_k(c)$ is finite by virtue of Theorem 13.1. But the number of $k$-cap heights is finite by hypothesis so that the sum

$$m_k = \sum_c m_k(c)$$

is finite. That the connectivity $p_k$ is finite follows at once from Corollary 12.1.

To establish the relations (13.3) let $N$ be an arbitrary positive integer. Let $e$ be a positive constant less than the cap spans of $k$-caps for which $k$ is at most $N$. That such a constant $e$ exists follows from the preceding lemma. Recalling that $m_k^e(c)$ is the $k^{\text{th}}$ $e$-type number of $c$ we see that for $k = 0, 1, \cdots, N$,

$$m_k^e(c) = m_k(c),$$

so that

$$m_k^e = m_k .$$

But if we set $\mu_k = m_k^e - p_k$ the relations (13.3) are satisfied in accordance with Corollary 12.3. Hence (13.3) holds for $\mu_k = m_k - p_k$ and for $n$ at most $N$. But $N$ is an arbitrary positive integer. Relations (13.3) thus hold as stated.

COROLLARY 13.2. *Under the conditions of the theorem*

$$m_k \geqq p_k , \qquad (k = 0, 1, \cdots)$$

*and for $k \geqq 0$,*

$$\mu_{k+1} + \mu_{k-1} \geqq \mu_k \qquad (\mu_{-1} = 0).$$

THE INSTITUTE FOR ADVANCED STUDY.

ANNALS OF MATHEMATICS
Vol. 41, No. 2, April, 1940

# THE GRAVITATIONAL EQUATIONS AND THE PROBLEM OF MOTION. II

BY A. EINSTEIN AND L. INFELD

(Received May 29, 1939)

## Introduction

The present paper contains a generalization and essential simplification of the theory concerning the problem of motion in the general Relativity theory as considered in I.[1] The method used there consisted of a distinction of some special coördinate system, the representation of matter by singularities and, finally, the use of a new method of approximation especially suited to the treatment of quasi-stationary fields. The change introduced here is: nothing is assumed in advance about the coördinate system except that it is galilean at infinity.

It turned out to be possible to develop the whole theory without any special equations expressing the choice of the coördinate system. The presentation given here is not only more general, but also essentially simpler than that given in I. The previous case, with its explicitly formulated coördinate conditions follows as a special case, distinguished by the natural character of the accepted coördinate conditions, from our more general considerations.

## 1. The Field Equations

We shall write down the equations of the gravitational field separating the linear expressions

$$(1,1)\quad \begin{aligned} &-\gamma_{mn|ss}+\gamma_{ms|ns}+\gamma_{ns|ms}-\delta_{mn}\gamma_{ls|ls}\\ &-\gamma_{0m|0n}-\gamma_{0n|0m}+2\delta_{mn}\gamma_{0s|0s}+\gamma_{mn|00}-\delta_{mn}\gamma_{00|00}+2\Lambda'_{mn}=0,\end{aligned}$$

$$(1,2)\quad -\gamma_{00|ss}+\gamma_{ls|ls}+2\Lambda'_{00}=0,$$

$$(1,3)\quad -\gamma_{0n|ss}+\gamma_{0s|sn}+\gamma_{ns|s0}-\gamma_{00|0n}+2\Lambda'_{0n}=0.$$

In writing down (1, 1)–(1, 3) we used the same notations and conventions as in I, that is:

Latin indices run from 1 to 3; repetition implies summation; strokes mean ordinary differentiation; $\gamma_{\mu\nu}$ (Greek indices running from 0 to 3) is:

$$\gamma_{\mu\nu}=h_{\mu\nu}-\tfrac{1}{2}\,\eta_{\mu\nu}\eta^{\sigma\rho}h_{\sigma\rho},$$

[1] Ann. of Math., Vol. 39, 1, 1938 p. 66.

where

$$\eta_{mn} = -\delta_{mn}\,, \qquad (\delta_{mn} \text{ are Kronecker symbols})$$

$$\eta_{0n} = 0; \qquad \eta_{00} = 1,$$

and

$$h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}\,.$$

The $\Lambda'_{\mu\nu}$ contain terms non-linear in $h_{\mu\nu}$. We do not write them out explicitly since we do not intend to make use of them. The whole calculation leading to (1, 1)–(1, 3) is a very slight modification of that in I (p. 68, 69). The only difference is that we do not assume, as in I, the coördinate condition

$$\gamma_{ls|s} = 0; \qquad \gamma_{0s|s} - \gamma_{00|0} = 0$$

and, therefore, (1, 1)–(1, 3) differ by some new additional linear expressions from the corresponding equations in I.

Let us rewrite (1, 1)–(1, 3) in the form

(1, 4) $$\Phi_{mn} + 2\Lambda_{mn} = 0,$$

(1, 5) $$\Phi_{00} + 2\Lambda_{00} = 0,$$

(1, 6) $$\Phi_{0m} + 2\Lambda_{0m} = 0,$$

where

(1, 7) $$\Phi_{mn} = -\gamma_{mn|ss} + \gamma_{ms|ns} + \gamma_{ns|ms} - \delta_{mn}\gamma_{ls|ls}\,,$$

(1, 8) $$\Phi_{00} = -\gamma_{00|ss} + \gamma_{ls|ls}\,,$$

(1, 9) $$\Phi_{0n} = -\gamma_{0n|ss} + \gamma_{0s|sn}\,,$$

and therefore,

(1, 10) $$2\Lambda_{mn} = -\gamma_{0m|0n} - \gamma_{0n|0m} + 2\delta_{mn}\gamma_{0s|0s} + \gamma_{mn|00} - \delta_{mn}\gamma_{00|00} + 2\Lambda'_{mn}\,,$$

(1, 11) $$2\Lambda_{00} = 2\Lambda'_{00}\,,$$

(1, 12) $$2\Lambda_{0n} = \gamma_{ns|s0} - \gamma_{00|n0} + 2\Lambda'_{0n}\,.$$

The reason for collecting some of the linear expressions in $\Phi_{\mu\nu}$ and combining others with the non linear $\Lambda'$ expressions will become apparent later.

## 2. Lemma

Let us consider a set of functions

$$F_{ab\ldots kl}$$

skew symmetric in the indices $k, l$, and arbitrary in all other indices. We form a surface integral:

(2, 1) $$\int_{(S)} F_{ab\ldots kl|l} \cos(x^k, \mathbf{N})\, dS$$

taken over an arbitrary closed surface $S$, which does not pass through a singularity of the field. Here $(x^k, \mathbf{N})$ denotes the "angle" between the direction $x^k$ and the "normal" to $S$.

Putting:

$$F_{ab\ldots 23} = A_1\,; \qquad F_{ab\ldots 31} = A_2\,; \qquad F_{ab\ldots 12} = A_3\,,$$

we can write (2, 1) in the form:

$$\int_{(S)} \mathrm{curl}_n\, \mathbf{A}\, dS, \tag{2, 2}$$

which vanishes identically as it can be changed by Stokes' theorem into a line integral around the rim of the surface; if the surface is closed the integral is of zero length. This result does not depend on whether the closed surface encloses a singularity or not. Hence we proved that (2, 1) vanishes.

### 3. The Equations of Motion

Here, as in I, we treat matter as a singularity in the field. Let us assume that there are $p$ bodies each represented by a point singularity. The space coördinates of each such singularity will be functions of the time alone. We may represent their positions at any time by means of their spatial coördinates $\overset{\kappa}{\xi}{}^m(x^0)$; $\kappa = 1 \cdots p$. The index $\kappa$ above $\xi^m$ refers to the $\kappa^{\text{th}}$ singularity.

Let us write the functions $\Phi_{\nu m}$ defined in (1, 7)–(1, 9) in the following way:

$$\Phi_{mk} = (-\gamma_{mk|l} + \gamma_{ml|k} - \delta_{mk}\gamma_{ls|s} + \delta_{ml}\gamma_{ks|s})_{|l} \tag{3, 1}$$

$$\Phi_{0k} = (-\gamma_{0k|l} + \gamma_{0l|k})_{|l}\,. \tag{3, 2}$$

The expressions in (3, 1)–(3, 2) placed inside the brackets are skewsymmetric in the indices $k, l$ and the lemma of §2 can, therefore, be applied. Therefore

$$\int^{\kappa} \Phi_{mk} \cos\,(x^k, \mathbf{N})\, dS = 0; \qquad \int^{\kappa} \Phi_{0k} \cos\,(x^k, \mathbf{N})\, dS = 0. \tag{3, 3}$$

Here, $\kappa$ above the integral sign denotes that the surface encloses the $\kappa^{\text{th}}$ singularity. From the last equations and from (1, 4)–(1, 6) follows:

$$2\Lambda_{mn|n} = 0; \qquad 2\Lambda_{0n|n} = 0 \tag{3, 4}$$

$$\int^{\kappa} 2\Lambda_{mn} \cos\,(x^n, \mathbf{N})\, dS = 0; \qquad \int^{\kappa} 2\Lambda_{0n} \cos\,(x^n, \mathbf{N})\, dS = 0. \tag{3, 5}$$

The equations (3, 4) indicate that the integrals (3, 5) can not depend on the shape of the surface as long as the surface encloses the same singularity. But equation (3, 5) gained by the application of our lemma states more: the surface integrals vanish.

Each of the surface integrals, being independent of the shape of the surface, can give only a relation between the coördinates of the singularities and their

time derivatives. In (3, 5) we have a set of $4p$ differential equations which we shall call *"the equations of motion of the p-particles."*

### 4. Application of the new approximation method

In the new approximation method, explained more fully in I, we introduce the auxiliary time coördinate $\tau = \lambda x^0$ and regard every field quantity as a function of $(\tau, x^1, x^2, x^3)$. We assume further that the derivatives of every field quantity with respect to $x^0$ are small compared with the space derivatives, or, in other words we assume that the derivatives with respect to $\tau$ and to the space coördinates are of the same order. We also assume the following expansion for the $\gamma$'s:

$$(4,1)\qquad \begin{cases} \gamma_{mn} = \underset{4}{\lambda^4\gamma_{mn}} + \underset{6}{\lambda^6\gamma_{mn}} + \underset{8}{\lambda^8\gamma_{mn}} + \cdots, \\ \gamma_{00} = \underset{2}{\lambda^2\gamma_{00}} + \underset{4}{\lambda^4\gamma_{00}} + \underset{6}{\lambda^6\gamma_{00}} + \cdots, \\ \gamma_{0n} = \underset{3}{\lambda^3\gamma_{0n}} + \underset{5}{\lambda^5\gamma_{0n}} + \underset{7}{\lambda^7\gamma_{0n}} + \cdots. \end{cases}$$

$l$ in $\lambda^l$ is the exponent and not the index. The numbers written beneath the $\gamma$'s show the power of $\lambda$ with which each is associated in the expansion. The differentiation with respect to $(\tau, x^1, x^2, x^3)$ is denoted by a comma so that e.g.

$$\gamma_{mn|s} = \gamma_{mn,s}\,, \quad \text{but} \quad \gamma_{mn|0} = \lambda\gamma_{mn,0}\,.$$

The "stroke differentiation" of a quantity with respect to zero can be replaced by the "comma differentiation" with respect to zero if the power of $\lambda$ with which this quantity is associated is simultaneously raised by one. To express this explicitly we use numbers under zeros, written after the comma, e.g.:

$$\underset{2l}{\lambda^{2l}}\gamma_{mn|0} = \underset{2l\ \ 1}{\lambda^{2l+1}}\gamma_{mn,0}, \quad \text{or} \quad \underset{2l}{\lambda^{2l}}\gamma_{mn|00} = \underset{2l\ \ 2}{\lambda^{2l+2}}\gamma_{mn,00}, \text{ etc.}$$

We can show further that our assumption (4, 1) induces a simple general rule of expansion which reads:[2]

Any component having an odd number of zero suffixes will have only odd powers of $\lambda$ in its expansion, while any component having an even number of such suffixes will involve only even powers of $\lambda$ in its expansion.

We want to find a solution of the field equations (1, 4)–(1, 6) by application of this approximation method. The obvious thing to do would be to split them according to our approximation method into the following system of equations:

$$(4,2)\qquad \begin{aligned} \underset{2l}{\Phi_{mn}} + \underset{2l}{2\Lambda_{mn}} &= 0 \\ \underset{2l}{\Phi_{00}} + \underset{2l}{2\Lambda_{00}} &= 0 \\ \underset{2l+1}{\Phi_{0n}} + \underset{2l+1}{2\Lambda_{0n}} &= 0. \end{aligned}$$

[2] I, p. 78.

But such a procedure is generally possible only when singularities are absent. In the case in which "the equations of motion" expressed in (3, 5) are not satisfied by arbitrary functions $\overset{\kappa}{\xi}{}^m(\tau)$ this procedure is not permissible. This can be seen from the following. We develop, say, the first equation (3, 5) with respect to the parameter $\lambda$:

$$\sum_{l=1}^{\infty} \lambda^{2l} \int \underset{2l}{2\Lambda_{mn}} \cos(x^l, \mathbf{N})\, dS = 0, \qquad \kappa = 1 \cdots p,$$

or putting

$$\int \underset{2l}{2\Lambda_{mn}} \cos(x^n, \mathbf{N})\, dS = 4\pi \overset{\kappa}{\underset{2l}{C_m}},$$

we have

$$\sum_{l=1}^{\infty} \lambda^{2l} \overset{\kappa}{\underset{2l}{C_m}} = 0, \tag{4, 3}$$

where $\overset{\kappa}{\underset{2l}{C_m}}$ depend on $\tau$ through $\overset{\kappa}{\xi}{}^m$ and their derivatives. Equations (4, 3) form a system of ordinary differential equations with $\lambda$ as a parameter the solution of which gives the motion of singularities. We must *not* conclude from (4, 3) that

$$\frac{1}{4\pi} \int \underset{2l}{2\Lambda_{mn}} \cos(x^n, \mathbf{N})\, dS = \overset{\kappa}{\underset{2l}{C_m}} = 0 \tag{4, 4}$$

because this may give us, in the general case, an infinite number of equations and it would be impossible to satisfy them by any system of functions $\overset{\kappa}{\xi}{}^m$. But the development of the field equations in the manner indicated in (4, 2) leads to the wrong equations (4, 4) just as the same argument which previously led to (3, 5) now leads to (4, 4). At each approximation step we may obtain different equations, inconsistent with each other, for the motion of singularities. We have, therefore, to reject the application of the approximation method which gives the equations (4, 2).

To avoid this difficulty we shall, therefore, proceed differently. Since this is an essential point we shall formulate the general idea of our considerations before giving the details. It consists of two steps:

1. Instead of the gravitational equations we shall introduce new equations which we shall call, for short, the "generalized equations." The gravitational equations are of the form

$$\Phi_{\mu\nu} + 2\Lambda_{\mu\nu} = 0, \tag{4, 5}$$

whereas the generalized equations are of the form

$$\Phi_{\mu\nu} + 2\Lambda_{\mu\nu} = C_{\mu\nu}, \tag{4, 6}$$

$C_{\mu\nu}$ being some specially chosen functions of $\tau$, $x^s$ and $\lambda$. The functions $C_{\mu\nu}$ have to be chosen in such a way that the system (4, 6) can be developed in the

parameter $\lambda$ and the straightforward approximation method applied without encountering the difficulties which have already been pointed out.

2. Since we solve in this way, by our approximation method, the generalized and not the gravitational equations we must be able to specialize our solution so that for a certain $\lambda$ we obtain the solution of our system (4, 5), that is, a solution of our gravitational equations.

Thus our method consists of generalizing (4, 5), using our approximation method and finally specializing our solution (by restricting $\lambda$) so that it fits (4, 5).

We shall state here without justification our choice of $C_{\mu\nu}$ ; this will be clearly justified later when we show that our approximation method can be carried out without any difficulties of a fundamental nature. We take

$$C_{mn} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{C}_m/\overset{\kappa}{r})_{,n} + (\overset{\kappa}{C}_n/\overset{\kappa}{r})_{,m} - \delta_{mn}(\overset{\kappa}{C}_s/\overset{\kappa}{r})_{,s}\}, \tag{4, 7}$$

$$C_{00} = -\sum_{\kappa=1}^{p} (\overset{\kappa}{C}_s/\overset{\kappa}{r})_{,s}, \tag{4, 8}$$

$$C_{0n} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{C}_0/\overset{\kappa}{r})_{,n} + (\overset{\kappa}{C}_n/\overset{\kappa}{r})_{,0}\}. \tag{4, 9}$$

On the right hand side we have $\overset{\kappa}{r}$ defined by

$$\overset{\kappa}{r}^2 = (x^1 - \overset{\kappa}{\xi}^1)^2 + (x^2 - \overset{\kappa}{\xi}^2)^2 + (x^3 - \overset{\kappa}{\xi}^3)^2,$$

that is the square of the "distance" of the point of the field from the $\kappa^{\text{th}}$ singularity. The $\overset{\kappa}{C}$'s are functions of $\tau$ and $\lambda$ which can be developed with respect to the parameter $\lambda$

$$\overset{\kappa}{C}_m = \sum_{l=1}^{\infty} \lambda^{2l} \overset{\kappa}{C}_{2l\,m}, \tag{4, 10}$$

$$\overset{\kappa}{C}_0 = \sum_{l=1}^{\infty} \lambda^{2l+1} \overset{\kappa}{C}_{2l+1\,0}. \tag{4, 11}$$

Now $\overset{\kappa}{C}_{2l\,m}$, $\overset{\kappa}{C}_{2l+1\,0}$ are functions of $\tau$ only which we shall determine later.

We can now split the equations (4, 6) and the use of our approximation method leads to the following system of equations:

$$\Phi_{2l\,mn} + 2\Lambda_{2l\,mn} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{C}_{2l\,m}/\overset{\kappa}{r})_{,n} + (\overset{\kappa}{C}_{2l\,n}/\overset{\kappa}{r})_{,m} - \delta_{mn}(\overset{\kappa}{C}_{2l\,s}/\overset{\kappa}{r})_{,s}\}, \tag{4, 12}$$

$$\Phi_{2l\,00} + 2\Lambda_{2l\,00} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{C}_{2l\,s}/\overset{\kappa}{r})_{,s}\}, \tag{4, 13}$$

$$\Phi_{2l+1\,0n} + 2\Lambda_{2l+1\,0n} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{C}_{2l+1\,0}/\overset{\kappa}{r})_{,n} + (\overset{\kappa}{C}_{2l\,n}/\overset{\kappa}{r})_{,0}\}. \tag{4, 14}$$

In discussing these equations let us begin with (4, 12). On the left hand side we have, as follows from (1, 7) and (1, 10):

$$(4,15)\qquad \underset{2l}{\Phi_{mn}} = -\underset{2l}{\gamma_{mn,ss}} + \underset{2l}{\gamma_{ms,ns}} + \underset{2l}{\gamma_{ns,ms}} - \delta_{mn}\underset{2l}{\gamma_{ls,ls}},$$

$$(4,16)\qquad 2\underset{2l}{\Lambda_{mn}} = -\underset{2l-1\,1}{\gamma_{0m,0n}} - \underset{2l-1\,1}{\gamma_{0n,0m}} + 2\delta_{mn}\underset{2l-1\,1}{\gamma_{0l,0l}} + \underset{2l-2\,2}{\gamma_{mn,00}} - \delta_{mn}\underset{2l-2\,2}{\gamma_{00,00}} + 2\underset{2l}{\Lambda'_{mn}}.$$

The unknown functions $\underset{2l}{\gamma_{mn}}$ which have to be calculated from (4, 12) are contained only in $\underset{2l}{\Phi_{mn}}$. All the $\gamma$'s in $\Lambda_{mn}$ are already known from the previous approximation steps as we see from (4, 16), remembering that $\Lambda'_{mn}$ does not contain linear expressions. This was the very reason for our division of the field equations into $\Phi$'s and $\Lambda$'s. We can now provide the lacking definition of $\overset{\kappa}{C}_m$. The $\underset{2l}{\overset{\kappa}{C}_m}$ are defined in such a way that the surface conditions are identically fulfilled in each approximation step and do not give restrictions for the motion. We shall show now that it is possible always to choose $\underset{2l}{\overset{\kappa}{C}_m}$ so that the motion is unrestricted. Forming our surface integrals we obtain, from (4, 12):

$$(4,17)\qquad \int^{\kappa} 2\underset{2l}{\Lambda_{mn}} \cos(x^n, \mathbf{N})\, dS = 4\pi \underset{2l}{\overset{\kappa}{C}_m}$$

as $\int^{\kappa} (1/r)_{,n} \cos(x^n, \mathbf{N})\, dS = -4\pi$, if we choose as the surface a small sphere enclosing the $\kappa^{\text{th}}$ singularity. But $\Lambda_{mn}$ can be determined from the previous approximation steps. The functions $\underset{2l}{\overset{\kappa}{C}_m}$ are, therefore, known if $\underset{l'}{\gamma}$'s are known for $l' < 2l$. In the equations (4, 12) the right hand side and $\Lambda_{mn}$ are known and we can calculate the $\underset{2l}{\gamma_{mn}}$ without encountering the difficulty with the surface integrals.

The equation (4, 13) written out explicitly according to (1, 8) and (1, 11) is

$$(4,18)\qquad \underset{2l}{\gamma_{00,ss}} = \underset{2l}{\gamma_{ls,ls}} + 2\underset{2l}{\Lambda'_{00}} + \sum_{\kappa=1}^{p} \left(\underset{2l}{\overset{\kappa}{C}_s}/r\right)_{,s},$$

where $\underset{2l}{\gamma_{ls}}$, $\underset{2l}{\Lambda'_{00}}$, $\underset{2l}{\overset{\kappa}{C}_s}$ are already known. If we exclude poles of a higher order in $\underset{2l}{\gamma_{00}}$ then there still remains the freedom of adding to $\underset{2l}{\gamma_{00}}$ a harmonic function of the type of a single pole. We shall return to this point later.

Finally there is the equation (4, 14) in which, as follows from (1, 9) and (1, 12)

$$(4,19)\qquad \underset{2l+1}{\Phi_{0n}} = -\underset{2l+1}{\gamma_{0n,ss}} + \underset{2l+1}{\gamma_{0s,ns}}$$

$$(4,20)\qquad 2\underset{2l+1}{\Lambda_{0n}} = -\underset{2l\ \ 1}{\gamma_{00,n0}} + \underset{2l\ \ 1}{\gamma_{ns,0s}} + 2\underset{2l+1}{\Lambda'_{0n}}$$

and $\overset{\kappa}{C}_{0}$ is determined, as was $\overset{\kappa}{C}_{m}$ previously, by taking the surface integral. Again the only unknown functions are $\gamma_{0n}$ in $\Phi_{0n}$.

In this case of generalized equations the surface integrals are, instead of (3, 5):

$$\int^{\kappa} 2\Lambda_{mn} \cos(x^n, \mathbf{N}) = 4\pi \overset{\kappa}{C}_{m} \tag{4, 21}$$

$$\int^{\kappa} 2\Lambda_{0n} \cos(x^n, \mathbf{N}) = 4\pi\{\overset{\kappa}{C}_{0} - \tfrac{1}{3}\overset{\kappa}{C}_{s}\overset{\kappa}{\xi}^{s}\}. \tag{4, 22}$$

To obtain a solution of our gravitational equations we must, therefore, assume

$$\left.\begin{aligned} \sum_{l=1}^{\infty} \lambda^{2l} \overset{\kappa}{C}_{2l\,m} &= \overset{\kappa}{C}_{m} = 0, \\ \sum_{l=1}^{\infty} \lambda^{2l+1} \overset{\kappa}{C}_{2l+1\,0} &= \overset{\kappa}{C}_{0} = 0, \end{aligned}\right\} \tag{4, 23}$$

and these are *the 4p equations of motion.*

In each approximation step we therefore proceed as though our aim were to solve not the gravitational equations (4, 5) but the more general equations (4, 6) which do not restrict the motion. Finishing our approximation procedure we go back to our gravitational equations by restricting the motion through (4, 23).

## 5. The condition $\overset{\kappa}{C}_{0} \equiv 0$

In (4, 23) we have $4p$ equations for the determination of $3p$ functions $\overset{\kappa}{\xi}^{m}(\tau)$. It can be shown that the motion is not over-determined since the $\overset{\kappa}{C}_{2l+1\,0}$ may be chosen arbitrarily and we may therefore assume

$$\overset{\kappa}{C}_{2l+1\,0} \equiv 0; \qquad \overset{\kappa}{C}_{0} \equiv 0. \tag{5, 1}$$

Assumption (5, 1) is consistent with the field equations and restricts the free addition of arbitrary poles to $\gamma_{00}$ mentioned in §4.

Replacing

$$\gamma_{00\,2l} \quad \text{by} \quad \gamma_{00\,2l} = \gamma'_{00\,2l} + \sum_{\kappa=1}^{p} \overset{\kappa}{\sigma}_{0\,2l}/r; \qquad \overset{\kappa}{\sigma}_{0} = \overset{\kappa}{\sigma}_{0}(\tau) \tag{5, 2}$$

will not change the equation (4, 18). Let us replace besides (5, 2)

$$\gamma_{0m\,2l+1} \quad \text{by} \quad \gamma_{0m\,2l+1} = \gamma'_{0m\,2l+1} - \sum_{\kappa=1}^{p} \overset{\kappa}{\sigma}_{0\,2l}/r\ \overset{\kappa}{\xi}^{m}_{1}. \tag{5, 3}$$

The only change which can be induced in equation (4, 14) is that coming from

$$(5,4)\qquad (\underset{2l+1}{\gamma_{0s,s}} - \underset{2l\ 1}{\gamma_{00,0}})_{,n}$$

as a comparison with (4, 19) and (4, 20) shows. The expression (5, 4) goes over by substitution of (5, 2) and (5, 3) into

$$(5,5)\qquad \underset{2l+1}{\gamma'_{0s,s}} - \underset{2l\ 1}{\gamma'_{00,0}} - \sum_{\kappa=1}^{p} (\overset{\kappa}{\underset{2l+1}{\sigma_0}}/r)_{,n}\,.$$

The comparison with (4, 14) shows that we can always assume $\overset{\kappa}{\underset{2l+1}{C_0}} \equiv 0$ by determining accordingly the functions $\overset{\kappa}{\underset{2l+1}{\sigma_0}}$ in (5, 2) and (5, 3).

The field equations and equations of motion are, therefore:

$$(5,6)\qquad \underset{2l}{\Phi_{mn}} + 2\underset{2l}{\Lambda_{mn}} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{\underset{2l}{C_m}}/r)_{,n} + (\overset{\kappa}{\underset{2l}{C_n}}/r)_{,m} - \delta_{mn}(\overset{\kappa}{\underset{2l}{C_s}}/r)_{,s}\}\,,$$

$$(5,7)\qquad \underset{2l}{\Phi_{00}} + 2\underset{2l}{\Lambda_{00}} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{\underset{2l}{C_s}}/r)_{,s}\}\,,$$

$$(5,8)\qquad \underset{2l+1}{\Phi_{0n}} + 2\underset{2l+1}{\Lambda_{0n}} = -\sum_{\kappa=1}^{p} \{(\overset{\kappa}{\underset{2l}{C_n}}/r)_{,0}\}\,,$$

$$(5,9)\qquad \overset{\kappa}{C_m} = \sum_{l=1}^{\infty} \lambda^{2l} \overset{\kappa}{\underset{2l}{C_m}} = 0\,,$$

representing the result of our considerations and the generalization of the theory developed in I.

We can, indeed, easily show that the field equations in I are a special case of (5, 6)–(5, 9).

Let us assume the coördinate conditions accepted in I:

$$(5,10)\qquad \underset{2l}{\gamma_{rs,s}} = -\sum_{\kappa=1}^{p} \overset{\kappa}{\underset{2l}{C_r}}/r$$

$$(5,11)\qquad \underset{2l+1}{\gamma_{0s,s}} - \underset{2l\ 1}{\gamma_{00,0}} = 0.$$

Because of (1, 7)–(1, 9) the field equations (5, 6)–(5, 8) take the form

$$(5,12)\qquad \underset{2l}{\gamma_{mn,ss}} = 2\underset{2l}{\Lambda_{mn}}\,,$$

$$(5,13)\qquad \underset{2l}{\gamma_{00,ss}} = 2\underset{2l}{\Lambda_{00}}\,,$$

$$(5,14)\qquad \underset{2l+1}{\gamma_{0m,ss}} = 2\underset{2l+1}{\Lambda_{0m}}\,,$$

and the surface integrals are

$$\int^{\kappa} \underset{2l}{2\Lambda_{mn}} \cos(x^n, \mathbf{N})\, dS = 4\pi \overset{\kappa}{\underset{2l}{C}}_m, \tag{5, 15}$$

$$\int^{\kappa} (\underset{2l+1}{2\Lambda'_{0n}} - \underset{2l}{\gamma_{00,n0}}) \cos(x^n, \mathbf{N})\, dS = 0, \tag{5, 16}$$

exactly as in I.

INSTITUTE FOR ADVANCED STUDY AND
UNIVERSITY OF TORONTO